AL-TR-1992-0145

AD-A271 319

2

ARMSTRONG LABORATORY

DEVELOPMENT OF THE UTC-PAB NORMATIVE DATABASE

Robert E. Schlegel
Kirby Gilliland

THE UNIVERSITY OF OKLAHOMA
1000 ASP, ROOM 314
NORMAN OK 73019

Mark S. Crabtree

LOGICON TECHNICAL SERVICES, INC.
P.O. BOX 317258
DAYTON OH 45431-7258

OCTOBER 1992

FINAL TECHNICAL REPORT FOR PERIOD JUNE 1990 - MAY 1992

Approved for public release; distribution is unlimited.

Headquarters
US Army Medical R&D Command
Fort Detrick, MD 21701-5012

AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-7022

NOTICES

TECHNICAL REVIEW AND APPROVAL

AL-TR-1992-0145

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Regulation 169-3.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Kenneth R. Boff

KENNETH R. BOFF, Chief
Human Engineering Division
Armstrong Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1992	Final, June 90 to May 1992

4. TITLE AND SUBTITLE

Development of the UTC-PAB
Normative Database

5. FUNDING NUMBERS

PE 62202F
PR 7184
TA 14
WU 23
F33615-88-D-0532

6. AUTHOR(S)

Robert E. Schlegel*, Kirby Gilliland*
Mark S. Crabtree

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

*University of Oklahoma
1000 Asp, Room 314
Norman OK 73019

Logicon Technical Services, Inc
P.O. Box 317258
Dayton OH 45431-7258

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory, Crew Systems Directorate,
Human Engineering Division, AFMC, HSC
Wright-Patterson AFB OH 45433-7022
Office of Military Performance Assessment Technology (OMPAT)
Walter Reed Army Institute of Research, Walter Reed Army
Medical Center, Washington DC 20307-5100

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AL-TR-1992-0145

11. SUPPLEMENTARY NOTES

Funded under OMPAT MIPR No. 90MM0501
*Subcontract to Southeastern Center for Electrical Engineering Education (SCEEE), St Cloud FL 34769. Work Order No. SCEEE-HER/90-0011.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report summarizes the development of a comprehensive normative database for a large subset of tasks from the Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB). Tasks were selected from the AGARD STRES, CTS, and Walter Reed batteries. Data were collected by the University of Oklahoma and Armstrong Laboratory. All data were analyzed at the University of Oklahoma to address issues related to task reliability, comparability of tasks across batteries, group vs. individual test administration, order of task presentation and battery sequence, test-retest time intervals, imposition of response deadlines, extended trial lengths, and the usefulness of psychometric state measures.

With few exceptions, the data showed remarkable consistency across task batteries and within task types. Task reliability varied as a function of the dependent measure. CTS data showed good correspondence to a previous large-scale CTS database. Task presentation order and battery sequence did not influence task performance. Response deadlines provided a faster mean response time, but at the expense of more missed responses. Extended trial lengths had a more profound effect on continuous motor tasks such as Unstable Tracking. Changes in the psychometric state measures of sleepiness and mood were logical reflections of time on task.

14. SUBJECT TERMS

Performance Assessment Batteries, Human Performance,
Cognitive and Psychomotor Performance

15. NUMBER OF PAGES

284

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std Z39-18
298-102

THIS PAGE LEFT BLANK INTENTIONALLY.

PREFACE

This report documents the work performed at the University of Oklahoma under SCEEE Subcontract HER/90-0011 for the Armstrong Laboratory, Performance Assessment and Interface Technology Branch (AL/CFHP) under contract F33615-88-D0532. It also presents the results of a parallel study conducted by AL/CFHP. The effort was sponsored by the Tri-Service Office of Military Performance Assessment Technology (OMPAT).

As outlined in the Statement of Work and the approved Project Design, an experimental study was conducted to provide normative data and a better understanding of a subset of tasks from the Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB). In addition to providing for the collection and summary of normative data for tasks from the AGARD Standardized Tests for Research with Environmental Stressors (STRES), the Criterion Task Set (CTS), and a subset of the Walter Reed Army Institute of Research Performance Assessment Battery (WRAIR PAB), the study examined issues related to task reliability, comparability of tasks across batteries, group vs. individual test administration, order of task presentation and battery sequence, test-retest time intervals, imposition of response deadlines, extended trial lengths, and the usefulness of psychometric state measures.

The list of people deserving attribution for a project as extensive as this is a long one. The authors gratefully acknowledge the contributions of graduate research assistants Randa L. Shehab, Scott H. Mills, Patrick L. Foster, and Ioannis Vasmatzidis and the work of the undergraduate support team (Mindy Mitchell, Rebecca Kempner, Tricia Baird, and Tammy Kasbaum) in collection, conversion, summarization, and analysis of the vast amounts of data. A special thanks goes to Mark S. Crabtree of Logicon Technical Services Inc. for his unselfish support of our efforts in addition to his collaborative input to the project design and his work as project coordinator and experimenter for the parallel effort at Armstrong Laboratory. Thanks also go to Gary B. Reid for his technical, financial, and personal support of our work over several years and for his insightful contributions during the design phase of this project. The authors thank Dr. Dennis L. Reeves for his role in the development of the UTC-PAB AGARD STRES battery and also for his contributions to the project design, and acknowledge the skilled and timely programming contributions of Kathy M. Winter, Sam J. LaCour, and Kathy Raynsford. The Armstrong Laboratory effort was greatly facilitated by Dr. Herbert Colle, Wright State University, who generously provided laboratory space and equipment, and expedited WSU approval of the research. Last but not least, the authors express their great appreciation to Dr. Frederick W. Hegge for his leadership role in the development of OMPAT, his contributions to the design of this project, and his funding support.

THIS PAGE LEFT BLANK INTENTIONALLY

SUMMARY

This report summarizes the development and analysis of a comprehensive normative database for a large subset of tasks from the Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB). The tasks were members of the AGARD Standardized Tests for Research with Environmental Stressors (STRES), the Criterion Task Set (CTS), and a subset of the Walter Reed Army Institute of Research Performance Assessment Battery (WRAIR PAB). Data were collected at the University of Oklahoma and in a parallel study conducted by Armstrong Laboratory. All data were analyzed at the University of Oklahoma to address issues related to task reliability, comparability of tasks across batteries, group vs. individual test administration, order of task presentation and battery sequence, test-retest time intervals, imposition of response deadlines, extended trial lengths, and the usefulness of psychometric state measures.

With few exceptions, the data showed remarkable consistency across task batteries and within task types. Task reliability varied primarily as a function of the dependent measure. CTS data showed good correspondence to a previous large-scale CTS database. Task presentation order and battery sequence did not influence task performance. Response deadlines provided a faster mean response time but at the expense of more missed responses. Extended trial lengths had a more profound effect on continuous motor tasks such as Unstable Tracking. Changes in the psychometric state measures of sleepiness and mood were logical reflections of time on task.

THIS PAGE LEFT BLANK INTENTIONALLY

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEVELOPMENT OF THE UTC-PAB NORMATIVE DATABASE

1.0 INTRODUCTION

The Tri-Service Working Group of the Office of Military Performance Assessment Technology (OMPAT) undertook a program to develop task batteries and a standardized test methodology for human performance assessment. One major part of this assessment effort has been the development of the Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB), a specialized human performance task battery for laboratory and field research. The UTC-PAB consists of numerous human performance tasks organized in various sub-batteries. One of the most recently developed and more sophisticated of these batteries is the UTC-PAB/AGARD STRES Battery (Reeves, Winter, LaCour, Winter, Vogel, and Grissett, 1990). Other UTC-PAB supported batteries include the U.S. Air Force Criterion Task Set (CTS; Shingledecker, 1984; Shingledecker, Acton, and Crabtree, 1983) and the Walter Reed Army Institute of Research Performance Assessment Battery (WRAIR PAB; Thorne, Genser, Sing, and Hegge, 1985).

The present project was initiated in response to the need for a better understanding of several tasks in the UTC-PAB, namely the subset comprising the STRES battery. In its earliest stages the project was conceived to address a fairly specific need, namely the development of a normative database for the STRES battery. This was expanded to focus on two other batteries that are closely associated with OMPAT efforts in battery standardization and development, the CTS and the WRAIR PAB. A fundamental objective of the effort was to develop an integrated database that could be used not only for normative data comparisons, but also to answer basic questions regarding how subjects respond on tasks implemented in a specific battery, and how task behavior varies across batteries with similar tasks.

As the project evolved, it was expanded to accommodate several basic research questions developed by the principal investigators. Based on extensive past experience with the Criterion Task Set (Schlegel and Gilliland, 1990) and more recent experience with the UTC-PAB/AGARD STRES battery (Baird, Kasbaum, and Schlegel, 1990), other basic problems were identified as logical candidates for investigation. First, the degree to which response deadlines influence the nature, speed, and distribution of subject responses was identified as an important problem. Second, trial length was believed to be an important variable in determining the nature of the performance obtained, yet very little is known about this variable in task battery construction. It was also believed that trial length analysis might provide a possible way to explore the dynamics of

task performance over time. Third, numerous additional questions arose regarding the reliability of performance on the selected tasks. And, finally, there was concern for the effects of the sequence in which tasks (or batteries) are presented to subjects and whether any carryover performance effects (either learning or fatigue) existed. For these reasons, response deadlines, trial length, task order and battery sequence, along with reliability were among the major foci of the current project.

This report summarizes the experimental design and methods used to develop the normative database and to address the outlined research questions. It also provides statistical summaries and analyses of the performance data. Section 2.0 (Background) provides a description of the development of the UTC-PAB. Specific information regarding the creation of OMPAT and the evolution of the STRES battery, the CTS, and the WRAIR PAB are presented. Section 3.0 (Establishing a UTC-PAB Normative Database) presents the specific research goals of the project. This section is followed by Section 4.0 (Project Design and Method) which provides an extensive overview of the methodology and procedures used in the project. Section 5.0 (Project Results) presents the normative data, as well as analyses of the additional research problems addressed in this project. Finally, Section 6.0 (Summary) provides a brief list of the major research findings of the project. Extensive appendices that provide additional detailed information about the project data complete the report.

2.0 BACKGROUND

2.1 Need for Performance Assessment Batteries

As advances in technology increase the complexity of various operational environments, it has become increasingly important to develop methods for assessing and predicting the nature and amount of workload associated with specific operator tasks. Greater demands are now placed on designers to include *a priori* evaluations of not only the amount, but also the type of work required in newly designed work environments. In addition, a renewed interest in the effects of various environmental stressors has served to promote the need for highly reliable and valid measures of human performance.

Developments in at least two areas have enabled this need to be addressed. First, many evolving theories of cognition and human performance have played important roles in defining both the theoretical and practical limits of mental work capacity (e.g., Broadbent, 1958; Kerr, 1973; Navon and Gopher, 1979; Norman, 1968; Sanders, 1979; Treisman, 1969; Wickens, 1980). Second, technological advances, especially microprocessor developments, have allowed more intricate levels of task modeling, more control over task presentation, and more accurate and enlarged data collection ability. The linking of cognitive theory developments with recent microprocessor improvements has led to the construction of a number of sophisticated human performance task batteries that provide considerable promise for advancing both theory and application in a variety of fields. However, some of these batteries appear to be more advantageous than others due to more sophisticated levels of implementation, ease of use, and greater linkage to both applied and theoretical research.

The Office of Military Performance Assessment Technology (OMPAT) has engaged in a program to develop a standardized test methodology and task battery for human performance assessment. Among OMPAT's numerous accomplishments thus far are two advances that bear directly on this project. First, OMPAT has established a superordinate pool of candidate human performance tasks identified as the Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB). This UTC-PAB pool of tasks is organized largely in subsets or batteries. Second, OMPAT has established the Performance Information Management System (PIMS) to serve as a clearinghouse for UTC-PAB databases. Thus, as research is conducted on various UTC-PAB tasks, data from these studies can be consolidated and disseminated through a central, organized framework.

This report documents a major research effort to explore three UTC-PAB subsets. Aside from contributing to the normative database on the included tasks, the project explored numerous

research questions regarding the interrelationships of the batteries and tasks. The project also examined how selected variables believed to influence human task performance actually affect performance on these tasks.

2.2 Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB)

As outlined by Englund, Reeves, Shingledecker, Thorne, Wilson and Hegge (1985, 1987; see also Perez, Masline, Ramsey, and Urban, 1987), the concept for the Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB) evolved from the Military Performance Working Group in 1983. This group proposed the UTC-PAB as a primary measurement instrument for the evaluation of cognitive performance within the framework of a larger multi-level biomedical drug evaluation program. These efforts gave rise to more detailed task specifications for the UTC-PAB through the actions of the Joint Working Group on Drug Dependent Degradation of Military Performance ($JWGD^3$ MILPERF - Task Area Group workshop in November 1984, at the Naval Medical Research Institute, Bethesda, Maryland). This joint working group was the predecessor of the Office of Military Performance Assessment Technology (OMPAT).

This period saw the development of several human performance task batteries such as the U.S. Air Force Criterion Task Set (CTS; Shingledecker, 1984), the Walter Reed Army Institute of Research Performance Assessment Battery (WRAIR PAB; Thorne, Genser, Sing, and Hegge, 1985), and others (e.g., Bittner, Carter, Kennedy, Harbeson, and Krause, 1984). A major contribution by OMPAT was to bring together the most theoretically representative and practically relevant tasks from these numerous sources into one standardized format (Hegge, Reeves, Poole, and Thorne, 1985; Englund et al., 1985, 1987). At that point in its development, the UTC-PAB represented a pool of approximately 25 human performance tasks that were believed to assess various stages of cognitive processing, as well as both selective and divided attention functions. The UTC-PAB was also envisioned as a dynamic task battery system that would evolve over time (presumably, as tasks were added, modified, or removed), and could be used flexibly by adopting a "core" subset of tasks, or by constructing unique subsets of tasks for project-specific purposes.

An important related development in the evolution of the UTC-PAB was a meeting of the NATO Advisory Group for Aerospace Research and Development (AGARD), Aerospace Medical Panel Working Group 12 on Human Performance Assessment Methods (AGARD, 1989). The focal point of this meeting was to address the need for a task battery to investigate the influence of environmental stressors on human performance. Working Group 12 was formed to review the relevant human task performance literature and to select a subset of tasks that might be optimally combined to provide the AGARD Standardized Tests for Research with Environmental Stressors

(AGARD STRES) Battery. As noted in their report, the panel concluded that their effort to design the AGARD STRES Battery could be considered an extension of the OMPAT UTC-PAB approach to battery construction. Although the AGARD group had a much more specific purpose in mind, they adopted OMPAT's general approach of identifying task subsets, and they also selected tasks that were, for the most part, already included in the OMPAT UTC-PAB. For example, OMPAT had already incorporated many of the tasks from the USAF Criterion Task Set. AGARD provided specific recommendations and parameters for task presentations, but not specific computer programs requiring specific computer equipment configurations. Thus, from a broader perspective, the AGARD STRES Battery can be viewed as a subset of the UTC-PAB that has been more highly defined on one hand, while being presented in a more "machine-independent" manner on the other.

In response to the AGARD recommendations, OMPAT supported an effort to construct an AGARD STRES implementation within the framework of the UTC-PAB. This battery has been officially designated the UTC-PAB/AGARD STRES Battery (Reeves, Winter, LaCour, Winter, Vogel, and Grissett, 1990). It will be referred to in this report as the "STRES Battery." Thus, the STRES Battery is the latest and perhaps the most sophisticated battery to emerge from the OMPAT UTC-PAB program. However, at the present time, the STRES Battery, the CTS and the WRAIR PAB together probably play the most prominent roles as applied task batteries within the UTC-PAB framework. It is also important to note that each of these primary batteries was designed to address a specific research application as outlined in the brief overviews that follow.

2.2.1 Standardized Tests for Research with Environmental Stressors (STRES)

According to the original AGARD report, the AGARD STRES Battery was designed to evaluate the effects of environmental stressors on selected aspects of cognitive performance. For this reason, the specific tasks recommended by the AGARD Working Group were chosen for their conformity to basic Human Performance Theory (AGARD, 1989). In other words, these tasks are typically short duration, repetitive, highly-structured information processing tasks with well-defined stimuli linked readily to simply structured responses.

As noted above, the UTC-PAB/AGARD STRES Battery (STRES Battery) is the latest battery to emerge from the broader OMPAT UTC-PAB effort (Reeves et al., 1990). Working assiduously within the OMPAT UTC-PAB guidelines, Reeves and his colleagues drew from the UTC-PAB task pool selected versions of tasks recommended by the AGARD report. These tasks were then combined in one subset, and programmed for a standardized computer hardware configuration compatible with previous UTC-PAB task implementations. Currently, the STRES Battery

includes seven tasks that conform to the original AGARD STRES Battery recommendations (AGARD, 1989), and now represents one of the most advanced and consensually supported batteries within OMPAT's UTC-PAB program. The tasks include: Reaction Time, Mathematical Processing, Memory Search, Spatial Processing, Unstable Tracking, Grammatical Reasoning, and an Unstable Tracking/Memory Search dual task. More detailed descriptions of the STRES Battery tasks are provided in Section 4.3.

2.2.2 Criterion Task Set (CTS)

The CTS is a human performance battery developed as a tool to facilitate the evaluation of mental workload metrics (Shingledecker, 1984; Shingledecker, Acton, and Crabtree, 1983). In this regard, the CTS was originally designed to provide a set of standardized loading tasks to evaluate the relative sensitivity, reliability, and intrusiveness of a variety of proposed behavioral, subjective, and physiological indices of workload. The CTS was thus designed as a set of "benchmark" tests with which project-specific workload measures could be calibrated or compared. Of course, in addition to this benchmark function, the CTS has also been used as a standardized task battery for human performance assessment.

Perhaps the most important feature of the CTS is the fact that it was one of the first human performance batteries to be based on current information processing theories (i.e., Multiple Resource Theory - Wickens, 1992; and Processing Stage Theory - Sternberg, 1969). According to these theories, human mental performance is dependent on a number of stages, information processing resources, and specific functions. The CTS model hypothesizes three primary stages of processing: perceptual input, central processing, and motor output. There are specific mental processing resources associated with the input mode (either visual or auditory), the type of coding during central processing (either spatial/imaginal or abstract/symbolic), and the mode of response output (either manual or vocal). Also, the central processing stage is further divided to emphasize memory/recall functions and elementary mental activities such as information manipulation, reasoning, and planning/scheduling.

This model was used to guide the selection of CTS tasks which would be representative of the range of human operator performance. This was accomplished by operationally defining each element in the model in terms of the task characteristics associated with the resources required by the element. For example, resources associated with the visual perceptual/input element were defined in terms of the task characteristics of stimulus discriminability and numerosity of display sources. These characteristics would be represented by tasks requiring simple detection as well as monitoring and scanning. Additionally, it was recognized that any task is likely to make demands

at all processing stages. Thus, when actually selecting a candidate task for a specific element of the model, such as visual perceptual/input, the loading demands on central processing and motor/output elements were minimized.

An additional feature of the CTS battery of tasks is that (except for Interval Production) three versions of each task are included to provide graded loading levels (i.e., an easy, moderately difficult and difficult version). In this manner, the CTS actually provides a task taxonomy for evaluating workload metric sensitivity along the dimensions of mental workload type (i.e., resource/stage) and workload level (i.e., difficulty). This feature of the CTS also allows investigation of the sensitivity of stressor effects on performance.

A wide range of tasks from the literature on cognitive and psychomotor performance was screened according to the resource theory outlined above. The screening process resulted in the selection of nine tasks for CTS Version 1.0. Initial parametric studies were completed to determine estimates of training time needed for each task, to determine task pacing rates, and to establish standard task loading levels. The standard loading levels were determined through comparison of post-asymptotic performance measures and were corroborated by subjective ratings of task difficulty and complexity (Shingledecker, 1984).

Detailed task descriptions and the results of the initial parametric studies are provided by Shingledecker (1984). Training performance and training requirements for the tasks when presented as a complete battery are given in Schlegel (1986). Results from a large scale normative data collection study are presented in Schlegel and Gilliland (1990). Based primarily on the results of the latter effort, several of the tasks were modified in the development of CTS Version 2.0.

CTS Version 2.0 consists of nine tasks including: Display Monitoring, Continuous Recognition, Memory Search, Linguistic Processing, Mathematical Processing, Spatial Processing, Grammatical Reasoning, Unstable Tracking, and Interval Production. In addition to the comprehensive normative database study (Schlegel and Gilliland, 1990), the CTS tasks as a whole or in part have been successfully used as dependent measures in numerous human performance studies in settings as wide-ranging as cockpit workload analyses, evaluations of vibrotactile helmet-mounted displays (Lambert, 1990; Schlegel and Gilliland, 1990), and the establishment of index levels of excessive workload (Schlegel and Gilliland, 1986; Schlegel, Schlegel and Gilliland, 1988). More detailed descriptions of the CTS tasks examined in this effort are provided in Section 4.3.

2.2.3 Walter Reed Performance Assessment Battery (WRAIR PAB)

The Walter Reed Army Institute of Research Performance Assessment Battery (WRAIR PAB) has been in development and use for a number of years (Thorne et al., 1985). Like the STRES Battery, the WRAIR PAB was designed primarily as a means for assessing the influence of treatment effects (stressors, drugs, etc.), especially within the context of investigations utilizing repeated measures. To accommodate this type of research, the WRAIR PAB emphasizes relatively brief tasks and provides nearly unlimited alternate "test forms" through the use of an automated configuration file system. Once constructed, the configuration file will allow the researcher to present the same or different task battery sequences automatically for any given set of project-specific needs.

The WRAIR PAB includes the following tasks: Encoding/Decoding, 2-Letter and 6-Letter Visual Search, 2-Column Addition, Logical Reasoning, Digit Recall, Serial Addition/Subtraction, Pattern Recognition, Wilkinson Serial Reaction Time, Choice Reaction Time, Time Wall, Interval Production, Manikin, Stroop, Code Substitution, Matching to Sample, Delayed Recall and a number of self-assessments of physical and mental states. More detailed descriptions of the WRAIR PAB tasks used in this project are provided in Section 4.3.

2.3 Need for a UTC-PAB Normative Database Study

The efforts by OMPAT outlined above have resulted in important advances both in developing standardized human performance task batteries and in communicating the results of task battery research. The STRES, CTS, and WRAIR PAB's have each provided unique solutions to major assessment problems, and the establishment of the Performance Information Management System has provided the mechanism for more effectively and efficiently sharing task battery databases.

The study reported here focused on basic research questions regarding issues in task battery administration and use. Prior to this effort, very little normative data existed for some of these batteries, such as the new STRES Battery. Even in cases where some data existed, such as the CTS database (Schlegel and Gilliland, 1990), these data needed to be updated and compared to data from new versions of the batteries. Also needed was a considerable amount of basic research regarding the reliability and validity of the task batteries, the degree to which these batteries relate to one another, and additional explorations of variables that may affect task battery performance more globally. The present study was designed to simultaneously address a number of these issues. The following section outlines in detail the nature of this research effort.

3.0 ESTABLISHING A UTC-PAB NORMATIVE DATABASE

In understanding any large-scale, multifaccted project of this nature, it is usually helpful to first gain a global perspective of the project, which then aids in a more complete comprehension of the specific research goals. So that the reader can more easily assimilate the more complex details of the design and methods found in Section 4.0, this section of the report begins with an overview of the origins of the research project, as well as the rationale and context of the project. Following this introduction are more detailed discussions of the purposes and goals of the project.

This project represents a large comprehensive research effort drawing upon the collaborative input of numerous individuals across a number of research laboratories. In its earliest stages, the project was conceived to address a fairly specific need, namely the development of a normative database for the UTC-PAB. As the project evolved, it was expanded to accommodate several basic research questions that were generated by the principle investigators and by researchers at OMPAT and the Performance Assessment and Interface Technology Branch at Armstrong Laboratory, Wright-Patterson AFB.

The present project began in response to the need for a better understanding of several tasks in the UTC-PAB, namely the subset comprising the STRES battery. The first version of the STRES Battery was completed in mid 1990 (see Reeves, Winter, LaCour, Winter, Vogel, and Grissett, 1990) and was used with reasonable success for various pilot research projects.[1] However, as is common with the first versions of test batteries, this early use of the STRES Battery suggested the need for a number of modifications prior to widespread release. Also, little was known regarding the data one might expect from the STRES Battery. Certainly these tasks had been used before in various laboratory settings, but how subjects would respond to them as implemented in the STRES Battery was unknown. This is not an uncommon problem. Few of the available task batteries have well-defined normative databases. Another OMPAT supported battery, the CTS, is an obvious exception (see Schlegel and Gilliland, 1990). However, even the CTS has undergone a revision since this early database was developed. Thus, one very basic need was for the development of a normative database, preferably one that could be used not only to determine how subjects respond on tasks implemented in a specific battery, but also to provide some capacity to cross-reference to similar tasks in other batteries. For this purpose, the present project focused on three batteries that are closely associated with OMPAT efforts in battery standardization and development: the STRES Battery, the CTS, and the WRAIR PAB.

---

[1] The authors would like to acknowledge the leadership role of Dr. Frederick W. Hegge and Dr. Dennis L. Reeves in the development of the UTC-PAB AGARD STRES battery, and the skilled programming contributions made by Ms. Kathy M. Winter, Mr. Sam J. LaCour, and Ms. Kathy Raynsford.

At this point, a number of other basic research questions began to emerge for human performance battery researchers. Clearly, there were attempts to increase the standardization of task batteries. Both the efforts of OMPAT (i.e., UTC-PAB) and those of the AGARD Working Group were vivid examples. However, even with concerted efforts such as these to standardize both the nature of the tasks in a battery and the administration procedures, there remains the simple problem that numerous batteries exist -- batteries that may share the same types of tasks, but may still differ in subtle yet important ways. Therefore, one question that must be addressed is whether versions of the same task implemented within two different batteries share similar response characteristics.

Based on extensive past experience with the Criterion Task Set (Schlegel and Gilliland, 1990) and more recent experience with the UTC-PAB/AGARD STRES battery (Baird, Kasbaum, and Schlegel, 1990), other basic problems have been identified as logical candidates for investigation. First, the degree to which deadline conditions affect the nature, speed, and distribution of subject responses is an important problem. There are similar concerns regarding performance differences as a function of trial length. Another issue concerns the sequence in which tasks are presented to subjects and whether any carryover effects (either learning or fatigue) affect performance. Numerous additional questions arise regarding the reliability of performance on the selected tasks. For these reasons, deadline conditions, trial length, and task sequence, along with reliability were among the major foci of the current project.

Finally, research on task performance can become costly both in terms of time and money because human performance studies such as these require subjects who are well-trained on the tasks. The cost of this training time adds to each individual study. One technique explored in this project is the establishment of a pool of subjects who, through their involvement in this initial study, are all well-trained on the various task batteries. By maintaining this pool of experienced subjects, selective future studies can be performed more economically.

As outlined above, this project was a comprehensive research effort aimed at providing several research groups with data of mutual interest. These data range from fairly basic normative data through fundamental reliability and validity data to more theory-driven experimental data. Each of the specific research goals of this project are presented in more detail below.

3.1 Normative Database Development

Computer implemented human performance tasks, such as those developed by OMPAT, are analogues of everyday tasks or contain the essential components of everyday tasks. They are tools used frequently by behavioral and medical researchers for personnel selection or for experimental

evaluation. In personnel selection, standardized batteries can be used to screen large numbers of applicants for selected abilities (e.g., pilot selection). The typical use of these tasks for experimental evaluations employs one or more tasks administered both under a control or baseline condition and under a treatment or experimental condition. Experimental conditions typically include factors that have the potential for influencing operator performance such as external factors (drugs, workload, time pressure, time of day), internal factors (fatigue, motivation, effort), and environmental stressors (heat, noise, vibration). These types of controlled experiments provide useful, cost-effective ways to assess the risks associated with numerous factors in the work environment.

However, the usefulness of task batteries for such personnel and experimental purposes is largely dependent on the quality of the tasks and the degree to which the response characteristics of the tasks are known. In order to enhance the investigator's ability to provide accurate placement or isolate treatment effects, it is necessary to know precisely how people perform on the tasks under laboratory baseline conditions. With the aid of a carefully developed normative database, other investigators can thoughtfully structure their testing conditions to better replicate those conditions used to generate the database. They can also determine whether their baseline data are within reasonable expectations, thereby "calibrating" their use of the task battery and providing reassurance that both their equipment and subjects are performing effectively. In addition, there are certain situations where pretest baseline data are difficult or impossible to collect. In such situations, the pre-existing database can serve as normative data to which experimental data can be compared.

Thus, one of the major goals of this project was to initiate the development of a normative database for UTC-PAB task batteries, especially the newest battery (the STRES Battery), and thereby provide some crucial data to support the OMPAT performance database clearinghouse, known as the Performance Information Management System (PIMS). To accomplish this goal, baseline data were collected on a number of tasks selected from the STRES Battery, the CTS, and the WRAIR PAB (Section 4.3).

3.2 Reliability of UTC-PAB Measures

This project focused upon a number of unknown characteristics of UTC-PAB performance. Central among these unknown characteristics is the reliability of the UTC-PAB tasks. Reliability generally refers to the degree to which inherent measurement error in a test is reduced, thereby rendering the specific measurement *repeatable* (Guilford, 1954; Nunnally, 1967). Thus, reliability addresses the degree to which a measurement is consistent, especially over time.

Traditionally, there have been three general approaches to the measurement of reliability: split-half, test-retest, and alternate form (Guilford, 1954). Each of these approaches provides distinctly different information about the repeatability of a measure. Split-half reliability techniques generally address issues related to the internal consistency of a test and provide little value for highly structured and repetitive human performance tasks. For the purposes of this project, split-half reliability is probably of least importance.

Test-retest reliability techniques are primarily concerned with the stability of a measure over time. In this approach, measurements are performed on two (or more) occasions and compared. High positive correlations between the measurements suggest that the psychological functions or abilities that are measured remained stable during the two (or more) testing sessions. Test-retest techniques thus provide a fairly convenient method for assessing the repeatability of a measure.

There are some problems in test-retest reliability assessment however. The stability of a measure can be affected by factors that would normally contribute to error variance. These factors would include the subject's health, fatigue, boredom, emotions, and other environmental factors such as temperature, lighting, humidity, etc. Also, the experience of the subject during the first exposure to the test and any learning during the interim period can change the nature of the subject's response strategy on the second testing. The greater the delay between testing periods, the greater the potential for these conditions to affect the test score.

Alternate form techniques have features of both split-half and test-retest approaches. These techniques incorporate at least two versions of the measure, versions that are assumed to have equal means and variances. When administration of both tests can be close together in time, the resulting data reflect both equivalence in test content and stability in performance. As the test-retest interval increases, the alternate form technique becomes vulnerable to the same problems affecting test-retest approaches, namely fluctuations in internal or external variables that affect test outcome. The alternate form technique thus can provide information about the equivalence of the "psychological-measurement content" across the test instruments, as well as information about stability (Guilford, 1954; see also Gulliksen, 1950; Thorndike, 1951).

Human performance testing presents some especially difficult problems in reliability assessment. First, while generally simple in design, many of the tasks utilized in human performance testing require some degree of practice to develop proficiency. Thus, comparison of initial trials with subsequent retest trials is usually subject to considerable influence by the factors mentioned above that can moderate test-retest reliability. As a result, initial trials often serve as practice trials to overcome "learning effects" and to provide more stable "baseline" performance in

later trials. Unfortunately, the amount of training is often not standardized and, in some cases, training is not sufficient or not feasible due to methodological considerations. Even after sufficient training, it is often the case that the interval between baseline testing sessions is sufficiently long that fluctuations in internal and external factors may be large enough to affect reliability estimates. For many tasks, there appears to be a continuous decline in reliability with increasing time between testing sessions. In fact, Guilford (1956) notes that some psychomotor tests may yield split-half reliabilities of 0.90 to 0.95, while one-year test-retest reliabilities only reach approximately 0.70.

This study was designed to provide test-retest reliability data over several time intervals. These include 30 minutes, 24 hours, 1 week (5 days), and 3 weeks (19 days).

3.3 Training Requirements for UTC-PAB

This study provided the opportunity to partially examine the learning rate and the training requirements for a variety of tasks. The STRES Battery tasks were administered ten times during the training sessions and the CTS and WRAIR PAB tasks were administered five times each. Some limited conclusions can be drawn from these data. However, the reader should keep in mind that the gross similarity between versions of the same task on both the CTS and STRES Batteries has the effect of potentially compounding the training effect. In other words, the training a subject received on the CTS tasks could have easily been transferred to the similar STRES Battery tasks, and vice versa. Nonetheless, the data from the training sessions of this study should provide valuable information about the training requirements of these batteries.

3.4 Comparison of Similar Tasks Across Batteries

Many of these task batteries share common tasks. In fact, the CTS and STRES Batteries share no fewer than five tasks, almost identical in nature. These tasks include: Mathematical Processing, Memory Search (Sternberg), Spatial Processing, Grammatical Reasoning, and Unstable Tracking. Even though these tasks are nearly (or even completely) conceptually identical, they often bear subtle yet crucial differences as a result of the manner in which they are programmed and presented visually. Additional differences in the way subjects respond to these tasks may be related to different modes of subject response, such as keyboard versus special response keypad. One simple, yet important, question this study was designed to answer was how these various versions of the same task, implemented in different batteries, compared to one another. Of particular concern were the five tasks shared by the CTS and STRES Batteries.

3.5 Comparison of Group vs. Individual Testing

In any human performance training situation, especially one leading to a large normative database, a number of training factors might influence the nature of the data collected. One factor of importance is whether the collection of training and baseline data was conducted under group or individual conditions.

Group training and baseline data collection generally involves greater initial cost in terms of acquiring multiple training/data collection stations. However, this cost is often outweighed by the ability to simultaneously train and collect data on larger numbers of subjects more efficiently. By contrast, training and collecting data from subjects individually requires less of an initial facilities investment, but requires a considerable investment in staff resources and time.

Typically, the decision to adopt one approach or the other is based on available equipment, laboratory resources, staff resources, and the time available to complete the project. Unfortunately, these factors do not address perhaps the most important question. Specifically, what is the effect on the subject's performance as a result of being trained and assessed in groups, as opposed to individual training/testing conditions? Group versus individual testing dynamics have been investigated. It is well-known that groups can have significant influences on the actions and attitudes of individuals (Asch, 1951, 1956; Myers, 1962), and vice versa (McGrath, 1962). In fact, the simple presence of others can have "social facilitation" effects that can increase or decrease performance (Zajonc, 1965; see also Bond and Titus, 1983; Guerin, 1986).

To address this question, portions of this study were conducted utilizing group and individual testing protocols. The data collected during the individual training/testing sessions included only training and baseline testing (i.e., no retest reliability, deadline, or trial length investigations were conducted). These data provide the opportunity to assess the influence of group versus individual training and testing conditions on performance of the target task batteries.

3.6 Effects of Task Order and Battery Sequence

The potential influence of the order in which experimental treatments (or tasks) are presented is well-known in the experimental psychology literature, as are the methods for addressing the problem methodologically (Myers, 1980; Underwood and Shaughnessy, 1975). How this phenomenon influences task battery performance is unclear. Presumably, order effects are just as potentially threatening in task battery research as they are in other areas of behavioral experimentation. However, there are cases in human performance research where randomizing task presentation order may be difficult or impossible. In fact, the AGARD STRES Battery

guidelines specify a fixed sequence in which the tasks should be presented, but it is not known to what extent performance may be affected by this sequence.

Very little research has been conducted that assesses this problem in the area of task battery research, probably because most researchers just assume it exists and routinely counterbalance task presentation order to control for it. However, a follow-on analysis of the Schlegel and Gilliland (1990) investigation of CTS performance suggests that counterbalancing may not always be necessary. In this analysis, the authors compared the responses of subjects who were presented the CTS in a different order than that presented to the subjects in the original study. The results of this analysis revealed no significant differences between the two groups on any of the major dependent measures of the task battery. While random presentation of tasks is generally a prudent strategy, it is also important to know *when* order effects are a problem in task battery administration and, *if* order effects are a problem, to what degree. This project was designed to assess the influence of various task presentation orders on performance. In addition, the study was also designed to assess the order of task battery presentation on performance.

3.7 Effects of Imposing Response Deadlines

The current tasks in the UTC-PAB are essentially self-paced, with no appreciable response deadlines. Version 1.0 of the CTS imposed response deadlines for most tasks. Data from pilot studies prior to two major data collection efforts (Schlegel and Shingledecker, 1985; Schlegel and Gilliland, 1990) pointed to the fact that some deadlines were very strict and resulted in subject response failures on an unusually high number of trials. For other tasks, the deadlines provided little or no incentive for faster responses. As a result, the experimental testing reported in Schlegel and Gilliland (1990) was conducted using the Training option of the tasks. This option provided 15-second deadlines for all discrete response tasks. Based on this decision and the data from Schlegel and Gilliland (1990), CTS Version 2.0 uses modified deadlines. Table 1 compares the response deadlines for Versions 1.0 and 2.0 of the CTS and for the UTC-PAB/AGARD STRES.

Table 1. Response Deadlines (seconds) for CTS and STRES Tasks.

Task	CTS V1.0	CTS V2.0	STRES
Grammatical Reasoning	6.5	15.0	15.0
Memory Search	2.0	3.0	5.0
Mathematical Processing	3.0	15.0	15.0
Spatial Processing	2.5	15.0	15.0

There is no doubt that the imposition of a response deadline affects a subject's reaction time and the percentage of response failures as a function of the strictness of the deadline. An important question is how much the subject's actual response strategy, as reflected by the frequency distribution of the reaction times, is affected.

A recent pilot study involving the Memory Search, Grammatical Reasoning, and Spatial Processing tasks from the UTC-PAB/AGARD STRES battery confirmed that a response deadline typically reduces the mean reaction time for a session. A more important result is the fact that the standard deviation of reaction times for correct responses is significantly reduced. This result holds even for moderate deadlines for which the percentage of response failures does not appreciably increase.

The presence of a response deadline appears to motivate a subject to respond faster while maintaining an acceptable level of accuracy. Deadlines that are too strict place an added time pressure (stressor) on the subject and may unduly impair performance. The importance of this effect on subject performance measures for the discrete response STRES tasks was investigated so that reasonable deadlines may be established, especially if they are helpful in reducing performance variability and motivating subjects without providing additional stress.

3.8 Effects of Extended Trial Length

The UTC-PAB and related task batteries have relatively fixed trial lengths and these trial lengths are generally short (e.g., three minutes). While short trial lengths have several advantages, especially ease of administration and efficiency, they may also represent a major problem. Specifically, it is unclear whether such a short testing epoch reasonably represents general performance. One short epoch, even after practice and baseline trials, may simply not be sufficient to capture the nature of more generalized trends in performance. Even a small sample of short epochs may be insufficient, especially if there have been temporary fluctuations in external variables or abilities as mentioned above.

One factor that may distort the accurate view of treatment effects on performance during short trial lengths is the temporary recruitment of abilities that takes place upon initiation or change in workload. Borrowing from Selye's (1976) now famous physiological theory of stress adaptation, when faced with increased stress, organisms enact compensatory mechanisms (i.e., recruitment processes) to cope with the increased demand. The organism resists (copes) for some period and then collapses at rates related to the magnitude of the stressor. Recent work in such areas as selective attention, work strategy, dual-task paradigms, and the multiple resources models of

cognitive processing suggests that similar recruitment processes operate at the cognitive (as well as the physiological) level.

It is conceivable that short trial lengths may be providing data during that period in which the person is making a special effort to recruit resources to perform the task. Thus, it is possible that a subject could recruit resources under the most trying circumstances to perform well for three minutes. Additionally, the initial moments of task performance (even after practice) could be the most unstable. Extending the trial length or sampling data beyond a set period of on-task time may allow for sampling during more stable periods of performance, or at least periods more reflective of general performance levels.

This line of logic has serious implications for the issue of task sensitivity, that is, the capability of the task to adequately measure treatment effects. If cognitive resource recruitment does take place, and if it takes place in a compensatory form in the first few minutes of task performance or following a stressor, then extending trial length would allow sampling of performance under conditions of generalized resistance (or acclimation) rather than during compensatory recruitment. The opening minutes of task performance, protected by recruitment processes, would provide the maximum ability to resist the effects of not only workload onset, but also external variables of interest such as environmental stressors and drug effects. Longer time on task would provide the opportunity to pass through the initial recruitment phase and assess the influence of variables on performance under conditions during resistance, acclimation, or exhaustion.

Some recent pilot data from the authors' laboratory suggest that trial length may be a significant task parameter. Subjects were given practice trials on the CTS Memory Search task and then, in counterbalanced order, a 3-minute trial and a 21-minute trial (actually 7, 3-minute trials in rapid succession). The results indicated that depending on the epoch of the 21-minute period, subject data varied greatly in comparison to the 3-minute data. While these data are very preliminary, they do suggest that an increased understanding of the effect of trial length is important, especially with regard to its implications for task sensitivity.

In this study, trial length was examined by repeating the administration of the standard 3-minute trials of the same task. Performance during 6-minute, 12-minute, and 24-minute testing sessions was examined and compared to data from the standard trial length.

3.9 Usefulness of Psychometric State Measures

There have been numerous attempts to measure psychological states psychometrically. In many cases these attempts have been useful for assessing the influence of a variety of psychological and physiological stressors. For example, the developers of the WRAIR PAB have included psychometric scales to assess various psychological states, especially in relation to drug and disease effects.

For the purposes of this study, both the Mood Scale II and the Stanford Sleepiness Scale from the WRAIR PAB were included in the testing protocol. It was believed that aside from the value of developing some normative data for these scales, they might be useful as convergent measurements of task demand or other factors such as cumulative workload demand across the test session. Thus, the use of these scales was an exploratory attempt to confirm or identify a relationship between performance changes or differences and concomitant changes in mood.

3.10 Software Analysis/Evaluation

The task batteries utilized in this project varied in the degree to which they were "field proven." During the course of this project, a number of issues or problems related to software, hardware, and such matters as instructional sets and data management were identified. Feedback was provided to the developers of the individual batteries. This feedback has resulted in corrections and modifications incorporated in updated versions of the various batteries.

4.0 PROJECT DESIGN AND METHOD

4.1 Project Design

As noted in the previous section, this was a comprehensive research project aimed at addressing a wide range of research needs from fairly basic normative data through fundamental reliability and validity data to theory-driven experimental data. To accomplish such a research effort, the project was designed so as to provide a high degree of structure, yet provide the flexibility to explore basic research questions. Thus, by its nature, the project required a complex design incorporating tradeoffs between competing research needs. For example, to ensure enough subjects for a stable normative data base, training on all tasks (across batteries) had to occur simultaneously. While this may raise some questions about skill transfer between the batteries during training, it was deemed that the baseline data were of more importance and that some contamination of the training data was acceptable as a tradeoff. Any serious limitations of the data set will be noted in the results section of this report.

Figure 1 presents an overview of the design and testing protocol for this project. Data for this project were collected under the direction of two research teams: (1) Dr. Robert E. Schlegel and Dr. Kirby Gilliland at the University of Oklahoma, and (2) Mr. Gary B. Reid and Mr. Mark S. Crabtree at Armstrong Laboratory, Wright-Patterson AFB. As noted in Figure 1, Orientation, Training, and Baseline testing (Week 2, Days 1 and 2) for establishing the UTC-PAB normative database and for assessing group vs. individual testing effects was conducted for all subjects at both locations. This phase of the project involved an orientation session followed by five days of training sessions (labeled "T" on Figure 1), plus two days of baseline sessions (B). All subjects completed the selected STRES, CTS, and WRAIR PAB tasks. More details regarding orientation and specific data collection procedures will be provided in the Experimental Procedure Section (see Section 4.6). Data addressing additional research questions such as one-week and three-week reliability (R3, R4), deadline effects (D) and extended trial lengths (E) were derived from subsequent testing sessions conducted only at the University of Oklahoma. These additional sessions involved testing on all tasks for the reliability test sessions. Data collection for the deadline and extended trial sessions was restricted to the STRES battery tasks.

All Armstrong Laboratory subjects were trained and tested individually. Subjects at the University of Oklahoma were tested in groups of eight. The basic testing protocol for Armstrong Laboratory subjects lasted seven days. The protocol for University of Oklahoma subjects required approximately five weeks. Two complete cycles of the five-week, University of Oklahoma testing protocol were needed to acquire data from the specified minimum number of subjects.

Figure 1. UTC-PAB Project Design.

	WEEK 1					WEEK 2					WEEK 3					WEEK 4					WEEK 5				
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 1	Day 2	Day 3	Day 4	Day 5	Day 1	Day 2	Day 3	Day 4	Day 5	Day 1	Day 2	Day 3	Day 4	Day 5	Day 1	Day 2	Day 3	Day 4	Day 5
ORIENTATION Prior to Week 1	T	T	T	T	T	B	B				R	R	D/E	D/E	D/E						R	R	D/E	D/E	D/E
						R2	R1				R3										R4				
Data Collected																									
STRES Battery	X	X	X	X	X	X	X				X	X	X	X	X						X	X	X	X	X
CTS Battery	X	X	X	X	X	X	X				X	X									X	X			
WRAIR Battery	X	X	X	X	X	X	X				X	X									X	X			

Data for 65 subjects collected at the University of Oklahoma

Data for 16 subjects collected at the USAF Armstrong Laboratory

NOTE:

T = Training days

B = Baseline data collection days

D/E = Deadline or Extended Trial data collection days

R1 = thirty-minute test-retest

R2 = one-day test-retest

R3 = one-week test-retest

R4 = three-week test-retest

4.2 Subjects

Subjects were recruited from the campuses of the University of Oklahoma and Wright State University, and the experimental procedures were implemented under the authorization of their respective Institutional Review Boards and in accordance with AFR 169-3. To control for possible performance variability due to gender, only male subjects were selected. Also, due to the verbal nature of many of the tasks and the instructional sets, only native English speaking subjects were recruited. Subjects were screened for gross hearing and visual impairments. Details of the screening process are presented in the Subject Recruitment, Screening, and Orientation Procedures Section (see Section 4.6.1). Table 2 provides general summary information regarding the two subject samples.

Table 2. Subject Characteristics.

		Group Administration (Oklahoma)	Individual Administration (Armstrong Lab)
Number	(N=)	64	15
	Deadline Study	(33)	-
	Extended Trials	(31)	-
Age	Mean	21.0	21.6
	Std. Dev.	3.2	3.1
	Range	18-36	18-27
Right Handed		59 (91%)	15 (100%)
Class	Freshman	21 (33%)	6 (37.5%)
	Sophomore	18 (28%)	6 (37.5%)
	Junior	14 (22%)	0 (0%)
	Senior	7 (11%)	4 (25%)
	Graduate	4 (6%)	0 (0%)
GPA	Mean	2.82	2.89
	Std. Dev.	0.59	0.51
	Range	1.60-3.84	2.20-4.00

All subjects were paid for their participation in the project. Because this was a multi-session experiment, a bonus system was used to increase motivation and completion rate. Subjects that successfully completed the study were given a bonus payment. Armstrong Laboratory subjects

received $80.00 for participating in the initial two-hour orientation and the subsequent seven, two-hour training and baseline sessions. These subjects were paid an additional $15.00 for attending an end-of-study debriefing session. University of Oklahoma subjects participated in five additional two-hour sessions. They were paid $4.00 per hour plus a $1.00 bonus for each hour if they successfully completed the study. Their bonus was $24.00 for a total of $120.00 for successful completion of the study.

4.3 Task Selection

4.3.1 Performance Tasks

Because the UTC-PAB/AGARD STRES battery was the primary focus of this project, the tasks included in this battery took a higher priority in selection. There was also a definite interest in including CTS tasks because the CTS is one of the few batteries for which there is an established database (Schlegel and Gilliland, 1990). Another important question was the degree of similarity between tasks common to both the STRES battery and the CTS. Therefore, tasks were selected to maximize the information gained regarding the STRES battery while at the same time affording a maximum level of information on the CTS tasks, as well as comparative information across these batteries.

Another UTC-PAB related battery frequently used for screening and selection is the WRAIR PAB. While there is less overall overlap between the WRAIR PAB and the batteries mentioned previously, there is some task overlap and there are some additional tasks that are unique to the WRAIR PAB and worthy of comparison. The tasks selected for inclusion in this project are listed in Table 3.

Five of the tasks that were examined in this study were implemented in very similar versions in the STRES and CTS batteries. They are (1) Grammatical Reasoning, (2) Mathematical Processing, (3) Memory Search, (4) Spatial Processing, and (5) Unstable Tracking. Each of these tasks is described in detail below. Additional information can be found in Shingledecker (1984), Englund, Reeves, Shingledecker, Thorne, Wilson, and Hegge (1987), and AGARD (1989).

Grammatical Reasoning. This task requires subjects to respond true or false to a pair of simple statements that describe the ordinal relationship of symbols (e.g., @ # *). For example, the subject is presented the following:

@ PRECEDES #

PRECEDES *

@ # *

Table 3. Task List.

Battery Comparison			
STRES	(code)	**CTS**	(code)
Grammatical Reasoning	(GRM)	Grammatical Reasoning	(GR)
Mathematical Processing	(MTH)	Mathematical Processing	(MP)
Sternberg - 2 Character	(STN2)		
Sternberg - 4 Character	(STN4)	Memory Search - 4 Character	(MS)
Spatial Processing	(SPA)	Spatial Processing	(SP)
Unstable Tracking	(TRK)	Unstable Tracking	(UT)
Supplemental Tasks			
WRAIR	(code)	**STRES**	(code)
Manikin Task	(MAN)	Reaction Time (6 Blocks)	(RCT)
Time Wall	(TIM)	Dual-Task Combination	(CBO)
Interval Production	(INT)		
Subject State Measures*			
Stanford Sleepiness Scale	(STA)	Mood II Scale	(MOO)
* collected before and after STRES battery runs			

The subject determines whether the first statement is true or false by examining the order of the symbols on the bottom line. The subject then determines whether the second statement is true or false. If both statements are true, or if both statements are false, the subject responds by pressing the "match" button. If one statement is true and the other is false, then the subject presses the "non-match" button. In this example, the subject's response would be MATCH.

There are 64 possible statement variations. Each statement is presented once before any statement is repeated during a trial. The subject would normally be required to respond about once every 3 to 5 seconds during the three-minute trial. Response accuracy and reaction time

are recorded on disk. Percent correct and mean response time for correct responses were presented to the subject upon task completion.

Mathematical Processing. In this task, simple problems involving multiple arithmetic operations are presented, one at a time, and the subject calculates whether the solution is less than or greater than 5. The subject is instructed to respond by pressing a key designated to indicate "less than" or "greater than." For example:

$$3 + 8 - 2 =$$

In this example, the answer is 9 and the subject would press the key designated to indicate "greater than 5." The subject may receive up to 50 presentations during a three-minute trial. Response accuracy and reaction time are recorded on disk. Percent correct and mean response time for correct responses were presented to the subject upon task completion.

Memory Search (Sternberg). In this task, subjects are required to memorize a set of either two or four letters. Then, as letters appear on the screen one at a time, the subject decides if each letter appearing on the screen is a member of the memorized set. The subject responds "yes" or "no" using assigned keys. Up to 100 letters may be presented during a three-minute trial. Response accuracy and reaction time are recorded on disk. Percent correct and mean response time for correct responses were presented to the subject upon task completion.

Spatial Processing. This task requires that the subject view a four-bar column chart (called the "target" stimulus) for 1 second. The bars are approximately .5 cm wide with .5 cm spacing between them, and their height varies from 1.0 to 6.0 cm. After the target stimulus disappears, a "comparison" stimulus appears that is rotated either 90 or 270 degrees from the original target position. The bar lengths of the comparison stimulus may be the same or different from the target stimulus. The subject responds using keys designated as "same" and "different." Fifty to sixty of these stimulus pairs may be presented in a three-minute trial. Response accuracy and reaction time are recorded on disk. Percent correct and mean response time for correct responses were presented to the subject upon task completion.

Unstable Tracking. This task presents to the subject a cursor moving horizontally on the screen. Depending on the computer system being used, a knob or a joystick is moved by the subject in order to keep the cursor centered on the screen. This task requires continuous subject control for the duration of a three-minute trial. The subject inputs required in this task are similar to those required by simple video games. Root mean square (RMS) tracking error

and control losses are recorded on disk for each second of task performance. Tracking RMS error and total edge violations (control losses) were presented to the subject upon task completion.

These five tasks are popular laboratory-based tasks used by psychologists, human factors specialists, and other behavioral scientists to explore fundamental properties of human performance. These tasks are commonly found in some combination in a number of task batteries. However, it should be remembered that even similar versions of such tasks within a family of batteries, such as the UTC-PAB group of batteries, can have large or subtle variations that may lead to overall performance score differences. For example, the STRES version of the Spatial Processing task presents bars that are non-filled and narrower in their general proportion compared to those in the CTS version. It may be the case that these graphical features of the column graph stimuli differentially affect performance. For this reason, this project was designed to provide comparisons of those tasks implemented in different batteries.

In addition to the preceding tasks that are common primarily to the STRES and CTS batteries, the Reaction Time Task and the Memory Search/Unstable Tracking Dual Task (COMBO) from the STRES Battery were included. Tasks from the WRAIR PAB that were also included were the: (1) Manikin task, (2) Time Wall task, and (3) Interval Production task. The Mood Scale II and the Stanford Sleepiness Scale, both from the WRAIR PAB, were also incorporated in the testing protocol.

Reaction Time. The Reaction Time task presents numbers from two to five on the left or right side of the screen. Some of the numbers are degraded in appearance, are temporally unpredictable, require multiple key presses, or require the subject to "switch hands" in terms of response mapping. The subject is instructed to press one of two keys to indicate the side of presentation on the screen and whether the number is a two or three versus a four or five. Reaction time and accuracy are recorded. Total session length for six different testing conditions lasts approximately 15 minutes. The word "error" was displayed on the screen following each incorrect response. Percent correct and mean response time were presented to the subject upon task completion.

Manikin Task. This task presents to the subject a male figure holding a green square in one hand and a red circle in the other hand. The objects are not always in the same hand. The manikin may be shown standing upright or upside down, and facing toward the subject or facing away from the subject. Encircling the manikin is either a red circular border or a green square border. For each stimulus presentation, the border signifies which manikin-held object

is the target figure. The subject must decide in which hand the manikin is holding the target figure, and then press a key to signify the right or left hand. Sixteen different stimulus figures are possible. Three complete sets of the 16 stimuli (i.e., a total of 48 stimuli) constituted a single trial lasting two to three minutes. Response accuracy and reaction time are recorded. Percent correct and mean response time were presented to the subject upon task completion. During orientation and training days 1 and 2, each response was followed by the presentation of a "C" for a correct response or an "E" for an error. After training day 2, the feedback following each response was no longer presented.

Time Wall. This task presents the subject with a small, red square at the top of the screen. The square drops at a constant rate. After traveling approximately two-thirds of the distance down the screen, the square is obscured by a red wall. At the bottom of the wall there is an apparent open space into which the falling square should eventually land. The subject presses a key when he or she thinks the square has had time to reach the open space. Thus, the subject's time prediction is the critical dependent measure. As soon as the subject responds, a new square appears at the top of the screen and the task is repeated. The task lasts less than two minutes for a total of ten squares. The mean of the estimated time intervals for the total travel of the square was presented to the subject upon task completion.

Interval Production. The Interval Production task simply requires that the subject press a specified key at regular intervals of approximately one second. On the screen, the subject sees a circle with a pointer much like a clock hand. When the subject presses the response key, the pointer advances 1/60th of the circle. The trial lasts until 60 responses have been made. The duration between key presses was recorded. Upon completion of the task, the subject was presented with the mean interval duration.

The five previously described tasks common to the STRES battery and CTS (plus the combined dual task variation), together with the four tasks described above, constituted the total task configuration administered in the training, baseline, and retest phases of this project. Each of these tasks yields a number of dependent measures such as mean and standard deviation of the response time for correct and incorrect responses, percent correct, etc. In fact, some of the tasks yield far more measures than can reasonably be evaluated. For the purposes of this project, dependent measures for summary presentation and analysis were restricted to a selected group of primary measures deemed most relevant. These dependent measures are listed in Table 4. The specific administration order of the tasks will be described below under Experimental Procedure (Section 4.6).

Table 4. Response Measures.

Category	Code	Description
STRES/CTS Discrete Tasks		
"Overall" for all stimuli	*xxMNO	Mean RT for Correct Responses
	xxSDO	Std. Dev. of RT for Correct Responses
	*xxPCO	Proportion of Correct Responses
	xxSTIMO	Number of Stimuli
"Positive" type stimuli	xxMNP	Mean RT for Correct Responses
	xxSDP	Std. Dev. of RT for Correct Responses
	xxPCP	Proportion of Correct Responses
"Negative" type stimuli	xxMNN	Mean RT for Correct Responses
	xxSDN	Std. Dev. of RT for Correct Responses
	xxPCN	Proportion of Correct Responses
STRES/CTS Unstable Tracking		
	*UTEV	Number of Edge Violations
	*UTRMS	Root Mean Square (RMS) Error
STRES Reaction Time		
	*RTMN	Mean RT for Correct Responses
	*RTSD	Std. Dev. of RT for Correct Responses
	*RTPC	Proportion of Correct Responses
WRAIR Manikin Task		
	*MANMNCR	Mean RT for Correct Responses
	*MANPC	Proportion of Correct Responses
WRAIR Time Wall/Interval Production		
	*TIMMN/INTMN	Mean Time Estimate/Interval
	*TIMSD/INTSD	Std. Dev. of Time

*Primary dependent measures used in analyses

4.3.2 Subjective Psychometric Scales

In addition to the performance tasks, two psychometric scales of psychological state were administered. Scales of this type are often used as simple dependent measures, as convergent measures, or as measures to verify the validity of a manipulation within psychological research. The two scales were selected from the WRAIR PAB. These two subjective tests were completed by each subject following each complete STRES battery administration. Both of these scales were presented on the computer monitor. Subjects responded by simply pressing keys corresponding to the appropriate subjective response. The scales used were the Mood Scale II and the Stanford Sleepiness Scale.

Mood Scale II. The Mood Scale II is a variation of the Profile of Mood States (POMS--McNair, Lorr, and Droppleman, 1971). The Mood Scale II has 36 items addressing the following six factors: Activity, Happiness, Depression, Anger, Fatigue, and Fear.

ILLUSTRATION of MOOD SCALE II

You will be given a list of words that people often use to describe how they feel followed by the numbers 1 to 3. These numbers represent the degree to which each word describes how you feel:

1 = "NOT AT ALL"

2 = "SOMEWHAT OR SLIGHTLY"

3 = "MOSTLY OR GENERALLY"

Indicate how each word applies to HOW YOU FEEL NOW, by pressing '1', '2' or '3'. (The following words were presented one at a time, and the subject responded following the presentation of each word.)

MISERABLE, UNEASY, INACTIVE, ENERGETIC, BLUE, GROUCHY, LIVELY, GOOD, MEAN, ANNOYED, DEPRESSED, ALARMED, INSECURE, WEARY, ALERT, LAZY, CONTENTED, CHEERFUL, SAD, DOWNCAST, SATISFIED, ANGRY, LOW, AFRAID, BURNED UP, DROWSY, CALM, IRRITATED, JITTERY, VIGOROUS, PLEASED, ACTIVE, HAPPY, STEADY, HOPELESS, SLUGGISH

Stanford Sleepiness Scale. The Stanford Sleepiness Scale is a scale designed to assess the level of sleepiness experienced by the subject. The subject simply responds to the scale by

selecting the statement that best describes the level of sleepiness at that moment.

ILLUSTRATION of the STANFORD SLEEPINESS SCALE

CHOOSE ONE OF THE SEVEN STATEMENTS BELOW WHICH BEST DESCRIBES YOUR PRESENT FEELING. HOW YOU FEEL RIGHT NOW.

1. Feeling active and vital; alert; wide awake.
2. Functioning at a high level, but not at peak; able to concentrate.
3. Relaxed; awake, responsive, but not at full alertness.
4. A little foggy; let down; not at peak.
5. Foggy; slowed down; beginning to lose interest in remaining awake.
6. Sleepy; woozy; prefer to be lying down; fighting sleep.
7. Almost in reverie; sleep onset soon; losing struggle to remain awake.

4.4 Hardware and Software Requirements

4.4.1 STRES Battery Requirements

Detailed hardware requirements for the UTC-PAB/AGARD STRES battery are found in the manual for the battery (see Reeves, Winter, LaCour, Winter, Vogel, and Grissett, 1990). Briefly, the STRES battery requires an IBM AT or compatible computer running at a minimum of 8 MHz with 640 Kb RAM. A 10 Mb hard disk is also required with at least one 5.25 inch floppy disk drive. At least CGA compatible video is required with a color monitor. The system also requires either a Systems Research Laboratories, Inc. (SRL) LabPak or Tecmar/SSI Labmaster multifunction data acquisition board. An analog joystick with an output voltage range of $\pm$ 5.0 VDC is also required. The software was written in the C programming language and compiled to produce executable task modules.

4.4.2 CTS Requirements

The CTS battery is implemented on the Commodore 64 microcomputer. The system requirements include: (1) Commodore 64 microcomputer, (2) Commodore 1541 disk drive (preferably two), (3) Epyx™ FastLoad™ cartridge (optional), (4) Commodore 1526 printer (or compatible), (4) two Commodore 1702 color monitors (or equivalent), and a custom keypad and rotary control device. Detailed information on the CTS hardware requirements can be found in Shingledecker (1984). The CTS software was written in Commodore BASIC with calls to various machine language routines. The BASIC source code was compiled to produce executable task files.

4.4.3 WRAIR PAB Requirements

Hardware specifications for the WRAIR PAB can be found in the WRAIR PAB manual (Thorne, 1990; see also Thorne, Genser, Sing, and Hegge, 1985). System requirements include: (1) at least an IBM or IBM compatible AT microprocessor with math co-processor, (2) two 360 Kb 5.25 inch floppy drives or a 20 Mb hard disk drive, (3) 640 Kb RAM, (4) IBM compatible bus with four unused slots, (5) two RS232 serial ports and one parallel port, (6) EGA (128 Kb; 640 X350) color graphics, and (7) SRL LabPak or other multifunction board. The software was written in the interpreted Microsoft BASIC programming language.

Table 5. Hardware and Software Configuration.

STRES, WRAIR	CTS
Zenith Z-248 PC	Commodore 64 Computer
Math Co-processor	2 Commodore 1541 Disk Drives
Internal Hard Drive (task software)	2 Commodore 1702 Monitors
360K Floppy Drive (subject data)	Four-Button Response Keypad
Zenith ZVM138 Monitor	Rotary Tracking Controller
SRL Labpak Board	Oklahoma - Bourns Potentiometer
Tracking Joystick	Armstrong - Allen-Bradley Potentiometer
(MS4M6676, OEM Controls Inc.)	Epyx™ Fastload™ Cartridge
NAMRL STRES Version 4.01 (JAN '91)	CTS Version 2.01A
WRAIR PAB Version 3.42 (MAR '90)	(Version 2.01 modified for automatic task
GWBASIC Version 2.18	sequencing, file naming, and data storage)

4.5 Testing Facilities

4.5.1 University of Oklahoma Facilities. All testing of University of Oklahoma subjects was conducted in a three-room suite in laboratory space allocated to the Department of Psychology. One room (approximately 13 ft. by 20 ft.) served as the microcomputer workstation site. Another room of approximately the same size served as a data reduction and project management office. The third room served as an auxiliary room for interviewing, orientation, and miscellaneous activities. All of these rooms represent modern laboratory space with centrally controlled heating and air conditioning. Temperature in the room was maintained at approximately

68 degrees Fahrenheit throughout the testing session. Lighting in the room was modified through the use of three 40W indirect incandescent lighting fixtures to reduce video screen glare.

Figure 2 presents the workstation configuration in the main testing room. Four STRES testing stations were located along one wall just to the right of the entrance. These testing stations were approximately 3.0 ft. wide and 3.0 ft. deep. The testing stations were divided by acoustic panels (3 inches in width). Keyboards and controllers were placed on tables at the testing stations positioned at a height of approximately 28 inches. Monitors were placed on 10-inch high shelves at the back of the table.

At the other end of the testing room, along the two alternate walls, were located the two WRAIR and two CTS testing stations. The dimensions of the WRAIR and CTS subject testing stations were of the same approximate dimensions as the STRES testing stations. Due to a lack of additional acoustic separation panels, these pairs of subject testing stations were divided by large cardboard panels and an experimenter control station (approximately 3 ft. wide).

The adjoining data reduction and project management room contained a complete Commodore 64 system for data reduction, an IBM compatible microcomputer for data reduction and transfer to the University IBM mainframe computer, and a terminal for data analysis on the mainframe computer.

4.5.2 Armstrong Laboratory Facilities

All Armstrong Laboratory tests were conducted in a laboratory of the Psychology Department at Wright State University. The room measured approximately 14 ft. by 10 ft. Temperature in the room varied from 62 to 72 degrees Fahrenheit. The room was illuminated by two, 4 ft., ceiling-mounted, 40W fluorescent fixtures.

Two testing stations were located at opposite ends of the room. No subject booths or enclosures were used because only one subject was tested at a time. There was one computer system at each subject station. One test station consisted of a Zenith Z-248 with a 20 Mb hard drive, a 360 Kb floppy drive, a Zenith EGA card, a Zenith ZVM-138 EGA color monitor, a math co-processor, DOS 3.10, and Zenith BIOS 3.12. The other station consisted of a Commodore 64 computer, an Epyx™ Fastload™ cartridge, two Commodore 1541 floppy disk drives, a Commodore 1702 color monitor, and the response keypad and tracking task controller normally used with the CTS (Acton and Crabtree, 1985). The computer systems were located on tables that measured 42 inches long by 30 inches deep, and were adjusted to a height of 28 inches. Seat height of the lightly padded, non-rolling, chairs was fixed at approximately 19 inches. The

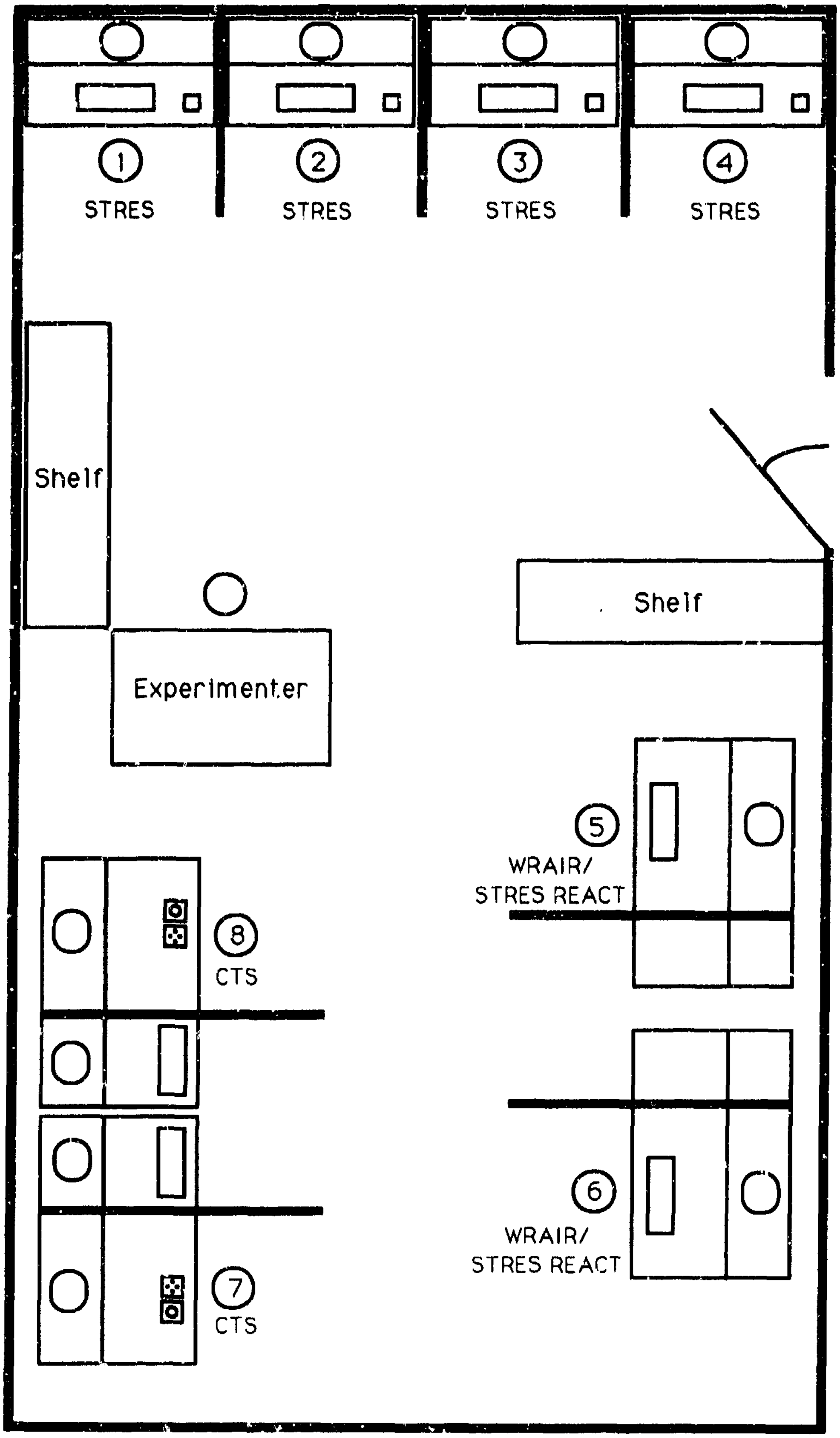


Figure 2. OU Test Facility Configuration.

experimenter was seated in a sound-reduction cubicle such that he could observe subject performance.

4.6 Experimental Procedure

4.6.1 Subject Recruitment, Screening, and Orientation Procedures

Subjects were recruited primarily from the undergraduate academic community at the University of Oklahoma and Wright State University. Following institutional review board approval of the project on each respective campus, the experimenters posted recruitment announcements and disseminated information about the project. Students who indicated an interest in the project either through bulletin board sign-up procedures or by contacting the experimenters were assigned an orientation appointment. The orientation appointment lasted two hours, was conducted individually, and consisted of screening procedures and orientation procedures.

Upon arrival for their appointment, subjects completed an informed consent form that described the general nature of the project and the expectations for their participation. The consent form also described the nature of the payment schedule including the rate of payment and the bonus agreement. All subjects who requested orientation interviews agreed to participate. Subjects then completed a biographical data sheet that included information such as name, local and permanent addresses, phone numbers, and appropriate times to contact them during the day. Subjects were questioned regarding any gross hearing or vision problems and the use of any medications, especially central nervous system stimulant or depressant medications. A brief visual acuity examination was conducted using a standard Snellen eye chart. This was to confirm that all subjects had normal or corrected vision of approximately 20/20, and certainly no worse than 20/30. No subjects were eliminated at either testing site during the screening process. This level of success in recruiting qualified subjects may have been due in part to the fact that the announcements used in recruiting subjects listed many of the required characteristics, such as native English speaking males with hearing and vision in normal ranges. Thus, subjects who sought orientation appointments were already self-selected based on these announced restrictions. Subjects were then given the opportunity to ask any remaining questions they had about the project, the payment procedures, or their commitment to participate. They were then scheduled for training and testing sessions. At this point, most subjects agreed to answer some ancillary questionnaires that required about 25 minutes to complete.

Following the screening procedures, subjects were provided an individual orientation to the tasks being evaluated in the project. Because so many performance tasks were involved in the project, it was believed that prior task description and instruction, and a brief exposure to each task

type would aid subjects in mastering and performing the task batteries. As a result, the first training day required much less task description and instruction, and more time was devoted to actual practice on the tasks.

Subjects were seated in front of a microcomputer, were given either an oral or written description of each task, were given an opportunity to ask questions and clarify their understanding of the task, and were then given an opportunity to perform a brief orientation trial of the task usually lasting no more than approximately one minute. As noted previously, the STRES and CTS batteries share similar versions of numerous tasks. Thus, subjects were not instructed on all tasks, but were informed that variations of certain tasks would be encountered during actual training and testing sessions. The subjects were provided orientation trials on the Grammatical Reasoning, Mathematical Processing, Memory Search, Spatial Processing, Unstable Tracking, and Reaction Time tasks from the STRES battery, as well as the Manikin task from the WRAIR PAB. These constituted the most complex tasks or those tasks with the most complex instructional sets. This orientation procedure lasted approximately one hour.

4.6.2 Features of the Testing Protocol

As noted in the Project Design section above, this project required a complex design to accomplish its many goals. Constructing the design included concern for such issues as achieving adequate training to obtain asymptotic performance on the various tasks, collecting baseline data at an optimal period in the project, addressing issues such as task sequence and battery sequence effects, scheduling retesting periods to obtain meaningful reliability data, and providing adequate testing time for the exploration of basic research questions. Reference to Figure 1 will reveal a number of unique design features constructed to meet these competing research needs.

Training, baseline, and additional data collection at the University of Oklahoma was conducted during the five weeks noted in Figure 1. An additional week prior to these sessions was needed for subject recruitment and orientation appointments as noted in Section 4.6.1 above. Thus, the project protocol required a complete data collection cycle of about six weeks. The laboratory configuration at the University of Oklahoma accommodated eight subjects during each two-hour training/test session. The four sessions scheduled each day allowed for the testing of 32 subjects per data collection cycle. Two complete data collection cycles were needed to acquire data from the desired number of subjects, i.e., a total of 64 subjects.

Armstrong Laboratory subjects required a shorter, two and one-half week data collection cycle because they were not involved in the reliability, deadline, and extended trial data collection. Because Armstrong Laboratory subjects were run individually, a maximum of four subjects in

two-hour sessions could be scheduled for any data collection cycle. Four cycles were needed to collect the desired number of Armstrong Laboratory subjects, i.e., a total of 16 subjects.

Data from the Armstrong Laboratory subjects during the training and baseline sessions, and the corresponding data from the University of Oklahoma subjects, provided the database for exploring training requirements on the task batteries, establishing the normative data for the various tasks, assessing fairly immediate test-retest reliability levels, and comparing the effects of group versus individual subject task administration. That the data were collected in different laboratories, with different laboratory personnel, and with different computer equipment (although the equipment models were the same), provided the opportunity to assess the robustness of the software, hardware and testing procedures to such subtle testing influences.

The additional data collected at the University of Oklahoma were designed to address other research questions. Repeated baseline testing sessions in Weeks 3 and 5 (labeled "R" on Figure 1), provided additional data for reliability analyses over longer time periods. Having more than one retest session during these weeks provided the opportunity for subjects to overcome any declines in performance efficiency that might have occurred due to the passage of time. By re-establishing baseline performance in this cost-efficient manner, the subjects were then fully capable of providing additional data. This situation was capitalized upon by following the repeat baseline data collection sessions with testing sessions aimed at answering questions of basic theoretical significance (i.e., questions addressing deadline and extended trial effects).

4.6.3 Training and Testing Procedures

Following subject recruitment, screening, and task orientation, the subjects completed one week (i.e., five consecutive days) of training sessions. The training sessions each lasted two hours. During the first session, each task was introduced by the experimenter and described. (Specific battery and task sequences are described in the next Section 4.6.4.). Because the subjects had prior orientation trials on most of the tasks, a simple description was all that was needed for the subjects to begin performing the task. Because some tasks were difficult or had complex instructional sets (e.g., Grammatical Reasoning, STRES Reaction Time task), the experimenter had to take additional care in presenting the task and close scrutiny was given each subject's performance to ensure correct understanding. Those tasks that were not included in the prior orientation had to be carefully presented for the first time, again with close scrutiny of the subject's performance. Subsequent training sessions required little additional instruction, although the experimenters were always scrutinizing the subject's performance to ensure understanding and compliance with task requirements.

Following this first week of training sessions, the subjects returned after the weekend for baseline sessions on the first two days of the second week. During these baseline sessions, the task testing sequences for each subject remained the same as those performed in the training sessions.

In the third week of each data collection cycle, half of the University of Oklahoma subjects (N=16) returned for two days of retest reliability testing. These data were collected five days following the last test session and are referred to as **"one-week" retest interval data** throughout this report. The task testing sequence for each subject was the same as that assigned during training and baseline sessions. The remaining three days of the third week were used for investigations of response deadlines (see Section 4.6.5) or extended trial lengths (see Section 4.6.6). On these days, subjects received trials with deadlines of varying lengths or with varying trial lengths imposed. In the case of varying trial lengths, fewer actual trials were possible.

No testing was conducted during the fourth week of either data collection cycle. During the fifth week of each data collection cycle, the remaining half of the subjects in that cycle (N=16) returned for retest reliability testing in the same manner as described previously for the one-week retest sessions. While these data were collected 19 days after the last previous test session, they are referred to as the **"three-week" retest interval data** throughout this report. As before, the three remaining days of the fifth week were used for additional investigations of deadline and extended trial effects. Deadline testing was conducted with the one-week retest subjects during the first cycle and with the three-week retest subjects during the second cycle. Extended trial testing was performed vice-versa.

At the conclusion of the two data collection cycles, 64 subjects from the University of Oklahoma and 16 Armstrong Laboratory subjects had been trained and tested on baseline conditions. Thirty-three subjects from the University of Oklahoma had been tested again at the one-week retest baseline period along with additional deadline and extended trial testing, and thirty-one University of Oklahoma subjects had been tested again at the three-week retest baseline period along with additional deadline and extended trial testing.

4.6.4 Battery and Task Sequences

Task sequence and battery sequence effects were assessed throughout the study. Four subgroups of approximately twenty subjects each were tested using different task sequences. This includes approximately sixteen subjects from the University of Oklahoma and four Armstrong Laboratory subjects.

During each of the training, baseline, and repeat baseline sessions, half of the subjects performed two trials of each STRES battery configuration during the first hour while the other half of the subjects performed one session of the CTS and WRAIR PAB tasks. The subjects switched workstations during the second hour so that each subject could complete all tasks. Within each hour, those subjects performing the CTS and WRAIR PAB configurations switched workstations half way through. Major orderings of the batteries (and specific task orderings within the batteries) were presented in counterbalanced fashion. Table 6 presents the counterbalanced sequences of the batteries. Subjects were randomly assigned to one of the four task battery sequences. Subjects were trained and tested according to their assigned battery sequence throughout the project. That is, a subject did not change sequences at different stages in the project.

To construct the battery sequences described in Table 6, the entire complement of tasks used in the project was divided into two major orderings. One of the major orderings began with the administration of the WRAIR Sleep/Mood scales followed by a trial of most of the STRES battery tasks performed in one of four fixed orderings (described below). After a short break, the subject performed the second trial on the STRES tasks in the same task order as before, and finished this major ordering by completing the WRAIR Sleep/Mood scales a second time. This major ordering took subjects approximately one hour to complete.

The other major task ordering consisted of the counterbalanced presentation of the CTS tasks (one trial each in one of four task orders described below) and the WRAIR PAB tasks along with the STRES Reaction Time task. The WRAIR Sleep/Mood scales were not administered during this major ordering. This major ordering required approximately one hour to complete.

These major orderings of the batteries were counterbalanced to produce the sequences presented in Table 6. For example, the first column in Table 6 presents one of the sequences consisting of the STRES major ordering followed by the CTS/WRAIR PAB major ordering. The exact number of subjects performing each sequence is noted under the column heading.

Embedded within the counterbalanced battery sequences were four sets of pseudo-random task orderings. Table 7 presents these various task orderings. It was not feasible to present all possible task orderings. Therefore, four task orderings were selected on a rational basis. The first was the ordering specified by the AGARD STRES manual (AGARD, 1989). The second ordering was the ordering used in the development of the CTS normative database (Schlegel and Gilliland, 1990). The remaining two task orderings were constructed through random selection with the only restriction being that they not resemble the other existing task orderings. For a particular subject, the same task order was used for both the CTS and STRES tasks. To achieve a combined balance

Table 6. Battery Sequences.

Time	STRES/CTS/WR (N = 17 + 4)	STRES/WR/CTS (N = 17 + 4)	CTS/WR/STRES (N = 14 + 4)	WR/CTS/STRES (N = 16 + 3)
0	WR Sleep/Mood II	WR Sleep/Mood II	CTS	WRAIR/ STRES REACT
30 min	STRES Session 1	STRES Session 1		
	STRES Session 2	STRES Session 2	WRAIR/ STRES REACT	CTS
60 min	WR Sleep/Mood II	WR Sleep/Mood II		
	CTS	WRAIR/ STRES REACT	WR Sleep/Mood II	WR Sleep/Mood II
90 min			STRES Session 1	STRES Session 1
	WRAIR/ STRES REACT	CTS	STRES Session 2	STRES Session 2
120 min			WR Sleep/Mood II	WR Sleep/Mood II

across the sixteen combinations of battery sequence and task order, the same number of subjects (four University of Oklahoma subjects and one Armstrong Laboratory subject) was assigned to each combination.

Table 7. Task Orderings.

1 (N = 17 + 4)		2 (N = 14 + 4)		3 (N = 17 + 3)		4 (N = 16 + 4)	
STRES	**CTS**	**STRES**	**CTS**	**STRES**	**CTS**	**STRES**	**CTS**
MTH	MP	STN2		TRK	UT	GRM	GR
STN2		STN4	MS	SPA	SP	MTH	MP
STN4	MS	GRM	GR	GRM	GR	TRK	UT
SPA	SP	MTH	MP	MTH	MP	SPA	SP
TRK	UT	TRK	UT	STN2		STN2	
GRM	GR	SPA	SP	STN4	MS	STN4	MS

4.6.5 Imposed Deadline Procedures

Thirty-three subjects participated in three days of testing under response deadline conditions. The purpose of the deadline testing was to determine if the added time pressure would be beneficial or detrimental to subject performance. Possible effects may include a reduced mean RT for correct responses, a smaller standard deviation in RT, and an increase in the number of missed responses (with a corresponding decrease in percentage correct). Of additional interest was identification of changes in the shape of the RT probability distribution and investigation of the ability to generate the unconstrained distribution from knowledge of the deadline distribution. Only the STRES versions of Grammatical Reasoning, Mathematical Processing, Memory Search (2- and 4-character) and Spatial Processing were used. The STRES COMBO task was also performed in place of the Unstable Tracking task in the subject's normally assigned task sequence. Based on a summary of the data from the second baseline session for the first cycle of the University of Oklahoma subjects and based on previous CTS standardization data, deadlines were established at approximately the mean RT plus one standard deviation (short deadline) and the mean RT plus two standard deviations (moderate deadline). Table 8 presents the specific deadlines for each task.

Table 8. Response Deadlines (msec).

STRES Task	Short Deadline	Moderate Deadline
Grammatical Reasoning	6500	8000
Mathematical Processing	2500	3000
Memory Search - 2 character	650	800
Memory Search - 4 character	750	1000
Spatial Processing	1250	1500

The subjects were assigned to one of two groups (16 subjects in the short deadline group, 17 subjects in the moderate deadline group) with an attempt made to achieve an overall performance balance between groups across all tasks. An outline of the three days of testing is provided in Table 9. On the first day (Wednesday), each subject performed the STRES battery four times, with the tasks arranged in the same sequence used by that subject for training and baseline. In the first session, response deadlines were used according to the group to which the subject was assigned (short or moderate). The deadline was indicated to the subject by the disappearance of the stimulus from the display. Subjects were encouraged to respond before the stimulus disappeared. Responses made after the stimulus was removed were not recorded. In the second session, subjects were given the same instruction set but the stimuli were not actually removed from the display (pseudo-deadline). This was done in order to record all responses with the subject believing he was working under deadline conditions. The task software did not allow removal of the stimulus at the deadline while continuing to record responses after the deadline. The last two sessions of the first day served as training sessions under the actual deadlines assigned to that subject (short or moderate).

Table 9. Schedule for Deadline Testing.

MONDAY	TUESDAY	WEDNESDAY DAY 1	THURSDAY DAY 2	FRIDAY DAY 3
Re-test	**Re-test**	*Sleep/Mood*	*Sleep/Mood*	*Sleep/Mood*
		Deadline Training	Deadline Training	Pseudo Deadlines
Session	**Session**	Pseudo Deadline	*(Counterbalanced)*	*(Counterbalanced)*
		Sleep/Mood	No Deadline	No Deadline
		Deadline Training	Short Deadline	Short Deadline
		Deadline Training	Moderate Deadline	Moderate Deadline
		Sleep/Mood	*Sleep/Mood*	*Sleep/Mood*
				Debriefing

On the second day (Thursday), four sessions were conducted. Following a training session under the subject's assigned deadline condition, each subject performed one session under no deadlines, short deadlines, and moderate deadlines. The order of deadline presentation was counterbalanced across subjects.

On the final day (Friday), three sessions were conducted, one each under no deadline, pseudo short deadlines, and pseudo moderate deadlines. Subject instructions provided the information as to the condition under which the subject should perform the session. The various subject instructions for all three days and for the different conditions are provided in Appendix A. Following the sessions, subjects were debriefed to get their impressions of the entire study.

4.6.6 Extended Trial Procedures

The extended trial portion of this research project addressed a circumscribed task battery administration issue with extensive implications for basic research. At an applied level, the extended trial study examined whether the data collected in the standard three-minute trial corresponded to data collected in longer trial periods. This is an important question if users of task batteries make generalizations to "real world" work performance from baseline data on three-minute tasks. These generalizations may be markedly in error. Such probable discrepancies may be due to the differences found between data collected during a short three-minute period, when a subject can recruit all available performance resources, as opposed to longer periods where conservation and regulation of resources is required. Thus, at a simplistic level, longer trial lengths should lead to poorer overall performance due to greater demands on resources.

At another level, this question strikes at the core of research in such areas as stress and adaptation. Theories of stress and adaptation may provide important cues in understanding the differences found between data collected at varying trial lengths. Perhaps even more important, by utilizing advances in task batteries and workload technology, such extended trial studies may play an important role in testing stress theories and developing new theories of adaptive responding. The present project provided the opportunity to investigate this domain at a rudimentary level. By collecting successive trials of data on the same task over varying periods of time, simple yet informative questions could be answered about the ability of subjects to respond consistently over time.

Only a limited opportunity for collecting such data existed within the testing time limits of this project. Thus, data could not be collected on all tasks. Preference was given to the STRES battery tasks, specifically those that could also be generalized to CTS data (i.e., Unstable Tracking, Memory Search, Grammatical Reasoning, Mathematical Processing, and Spatial Processing). The

highest priority was given to the Unstable Tracking task due to the continuous nature of this task. The other tasks present the subject with discrete problems that are separated by short periods of time that could be used as intermittent rest periods. The Unstable Tracking task is continuous, thus providing a constant demand on resources.

Unfortunately, the existing task software would not permit extending the discrete trial length beyond the standard three-minute period. To produce extended trial lengths, multiple three-minute trials were administered in rapid succession minimizing any rest period between trials. The extended trial lengths selected included multiple three-minute epochs totaling 6-minute, 12-minute, and 24-minute trials. Figure 3 presents these extended trials. It can be seen from Figure 3 that various comparisons of the three-minute trial epochs provide a rudimentary means of answering numerous questions. For example, the combinations labeled "C" represent the comparisons of the Baseline three-minute trial with the average of the three-minute epochs for each of the longer trial lengths. At a fundamental level these comparisons can answer the question of whether overall performance during these trial lengths is equivalent. Comparisons labeled "A" provide an assessment of whether the first three minutes of performance during the varying trial lengths produce similar levels of performance. If subjects regulate their performance resources to match the time demands, these first three-minute epochs may not yield comparable performance levels. Comparisons labeled "B" simply provide an assessment of any differences between the first epoch and every other epoch in a 24-minute trial. Such a comparison yields information regarding the stability of performance across the total trial length. This type of comparison can be performed for the other multi-epoch trial lengths. Similar comparisons can be made with the three-minute baseline trial.

During the weeks of retesting (weeks three and five), the first two days of each week were dedicated to retest data collection. The remaining days were dedicated to deadline or extended trial data collection. Due to the limited amount of time for such data collection and the constraints due to each two-hour test session, the presentation of all counterbalanced combinations of extended trial length was not possible. As a result, the 24-minute trial length condition was given priority because it was the condition that placed the greatest degree of demand on the subjects and provided the longest amount of time on task to observe any differences that might occur.

Table 10 presents one of the four testing protocols for the extended trial length sessions. Not unlike the technique used for presenting the battery sequences, major orderings of trial lengths were constructed. The orderings each lasted approximately one hour so that two of them could be accommodated within the overall two-hour test session time constraint. One group of orderings included two 24-minute trials (never the same task). Because these 24-minute trials were expected

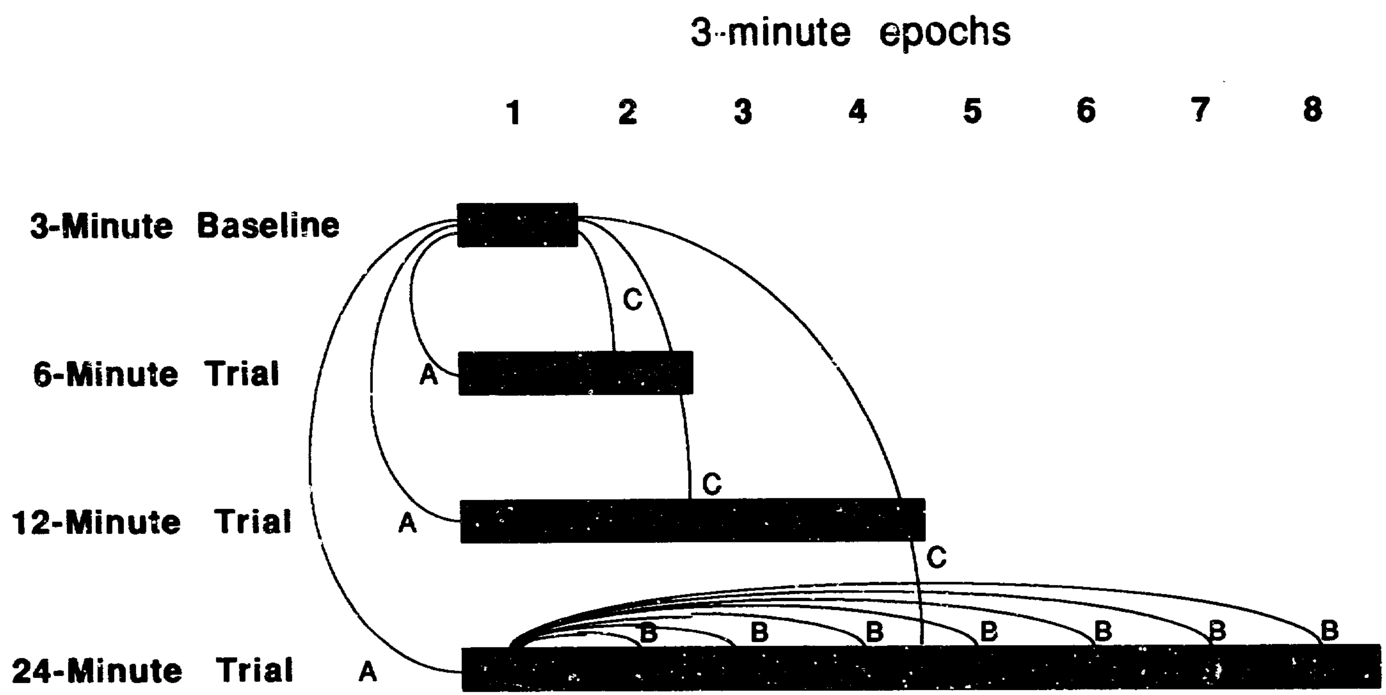


Figure 3. Extended Trial Analyses.

Table 10. Schedule for Extended Trial Study - Group A2.

MONDAY	TUESDAY	WEDNESDAY DAY 1	THURSDAY DAY 2	FRIDAY DAY 3
Re-test	Re-test	*Sleep/Mood*	*Sleep/Mood*	*Sleep/Mood*
		24-minute Tracking	24-minute Gram Reas	24-minute Spatial
		Sleep/Mood	*Sleep/Mood*	*Sleep/Mood*
		24-minute Sternberg	24-minute Combo	24-minute Math
Re-test	Re-test	*Sleep/Mood*	*Sleep/Mood*	*Sleep/Mood*
		12-minute Spatial	12-minute Gram	12-minute Tracking
		6-minute Combo	6-minute Combo	12-minute Sternberg
		6-minute Math	6-minute Gram	
		6-minute Combo	6-minute Spat	6-minute Tracking
		12-minute Math	12-minute Combo	3-min Combo
				Sleep/Mood
				Debriefing
		3-min Combo	6-minute Sternberg	
		Sleep/Mood		
			Sleep/Mood	

to be rather fatiguing, no more than two such trials were administered on any test day. This meant that the second group of major orderings included combinations of 6-minute and 12-minute trials. Unfortunately, due to technical constraints, it was never possible to administer the 6-minute trials first within this group of orderings, and thus within any test session.

On any specific test day, subjects received in counterbalanced order a combination of the two major groupings. Following the example in Table 10, during the first hour of Wednesday's test session, subjects were administered a 24-minute trial of the Unstable Tracking task followed by a 24-minute trial of the Sternberg task. During the second hour, the subjects received, in order, a 12-minute trial of Spatial Processing, 6 minutes of the COMBO task, 6 minutes of Mathematical Processing, 6 minutes of the COMBO task, 12 minutes of Mathematical Processing and a standard 3-minute trial of the COMBO task. As shown in the table, four sets of Sleep/Mood data were obtained during the two-hour period. Similar schedules were used for Thursday and Friday. Subjects were assigned to one of four extended trial schedules to achieve a balance among the groups based on overall baseline performance. All four schedules are provided in Appendix B.

The specialized testing periods dedicated to deadline and extended trial length also provided one of the few opportunities to collect data on other tasks of interest. For that reason, subjects were also administered the STRES COMBO task (a dual-task combination of Memory Search with four characters and Unstable Tracking). This task was simply included in the task administration protocol on these test days. Although performing these two tasks in combination was new to the subjects, the tasks themselves were well-practiced at this point. Thus, the introduction of the COMBO task at this point in the study was not viewed as disruptive in any sense.

4.6.7 Debriefing Procedures

Following the completion of training and test sessions, subjects participated in a debriefing session. These debriefing sessions lasted approximately 30 minutes and consisted of an oral interview with the experimenter(s). Table 11 presents the general contents of the debriefing interview. During these interviews, the experimenters attempted to determine the subjects' general response to the study and specific information about the tasks and batteries. Experimenters questioned the subjects regarding the nature of specific tasks, as well as comparisons between tasks. Subjects were also asked to provide information on strategies used to perform the tasks and points at which they may have changed those strategies. Hardware and software were evaluated, as were lab accommodations, initial instructions and staff support. Finally, those subjects that participated in additional trials investigating deadline and extended trial effects were asked to comment on that experience.

Table 11. Subject Debriefing Topics.

Overall Impressions
Most Preferred/Least Preferred Tasks
Strategies to Improve Performance ("tricks")
Grammatical Reasoning Symbols (40% of subjects)
Mathematical Processing Digits
Spatial Processing Bar Heights
Stimulus Sequence Memorization (STRES Spatial)
Comparison of CTS vs. STRES Hardware and Software
Comments on the Sleep and Mood Scales
Departure from Standard Instructions (e.g., finger placement)
Opinions on Response Deadlines and Extended Trials
Adequacy of Orientation Session
Treatment by Experimental Staff

5.0 PROJECT RESULTS

This chapter of the report presents the data and analyses from this project beginning with a description of the magnitude of the data collection effort and some characteristics of the database. A brief discussion of the results is included within each section. The first major section of this chapter presents baseline and training data for each of the task batteries. The presentation format is of a summary nature that should be particularly useful to those researchers interested in normative response patterns for each task. In the body of the report, emphasis is placed on measures typically of interest to researchers (i.e., mean and standard deviation). Also included for the convenience of the reader are tables with selected percentile groupings which allow classification of subjects into performance categories. Graphs for each major task measure and other detailed information are included in numerous appendices.

Following the first section on baseline and training data are sections that present the results of other analyses including task measure reliability, comparisons across task batteries, group versus individual testing procedures, task order and battery sequence effects, effects of deadline conditions, effects of extended trial length, and the usefulness of the psychometric state measures.

It is important to note here that the analysis of the individual versus group training effect (i.e., the comparison of University of Oklahoma and Armstrong Laboratory subjects) yielded no significant differences of any importance. This analysis will be discussed in more detail in Section 5.4. As a result of this finding, the data from subjects at the University of Oklahoma and the Armstrong Laboratory were combined. The data representing training and baseline performance reflect the total sample of 79 subjects with one major exception, the CTS Unstable Tracking task. The major summaries and analyses for this task include only the CTS Unstable Tracking data from the University of Oklahoma (see Section 5.1.3).

5.1 General Normative Database

This project involved the collection of a massive data base. Only a portion of those data is summarized within this report. Table 4 presented a list of the primary performance measures collected and analyzed and Table 12 presents a summary of the data collection effort. Over 20,000 data observations (subjects x trials x tasks), each containing numerous dependent measures, were collected and analyzed. More than 140 dependent measures were obtained on multiple days, and over 50 of these were included in some phase of the analysis. It is noteworthy that of the 20,000 plus observations, only seven were lost due to equipment or procedural errors. An additional 100 of the 20,000 represented outliers that were removed prior to the summaries and analyses. The majority of the deleted observations were due to identifiable subject errors.

Table 12. Summary of Data Collection Effort.

More than 140 Dependent Measures
More than 50 Dependent Measures Analyzed
Five Training Days/Two Baseline Days/Two Retest Days (all subjects) STRES - two sessions per day CTS, RCT, WRAIR PAB - one session per day
Three Days of Response Deadline Testing (33 subjects) STRES - four sessions per day (including CBO Dual-Task)
Three Days of Extended Trial Testing (31 subjects) STRES - 6, 12, and 24 minutes of GRM, MTH, STN4, SPA, CBO
More than 20,000 observations
Seven lost observations due to hardware/procedural errors
One Oklahoma and one Armstrong Lab subject removed due to poor motivation
Approximately 100 (< 0.5%) outlier observations removed due to subject errors

This project required subjects to return to the respective laboratories each day for two hours, on multiple days, across multiple weeks, to perform the same sequence of fairly repetitive performance tasks. In general, the subjects participating in this project understood the commitment in time and effort that they were making and understood the value of their contribution to this research. This was, in most cases, sufficient to maintain adequate motivation in the participants. However, the data from two subjects (one from each testing site) had to be eliminated due to poor motivation. These subjects exhibited chronic tardiness, missed sessions, and provided highly variable data. While it was clear that on any trial they were capable of providing data within the typical range of the other participants, these two subjects were repeatedly uncooperative and often provided data that could easily be characterized as "outlier" in nature. For these reasons, the data from these two subjects were eliminated from the analysis.

To facilitate the screening of outliers, the SAS Univariate Procedure was used separately on the University of Oklahoma and the Armstrong Laboratory data sets. In addition to providing the mean, standard deviation, upper quartile and lower quartile values, and a stem-and-leaf or histogram plot, the procedure presents a box-and-whisker plot and identifies the five lowest and the five highest data values. Through the box-and-whisker plot, SAS distinguishes between extreme data values that are 1.5 to 3 interquartile ranges away from the nearest quartile point vs. those that are more than 3 interquartile ranges away.

For each task in each battery, a listing was made of those subjects for whom potential outlier trials existed. A potential outlier trial was one which was placed in either extreme category on the box-and-whisker plot for response time (usually a long mean RT) or percentage correct (usually a low PC). A separate identification was made of those trials which differed from the mean by more than 4 standard deviations.

Each of these trials was closely examined to determine if the performance was consistent with that subject's typical performance. In a few cases, a particular subject was identified as a poor performer, either in general (as with the two eliminated subjects mentioned earlier) or for a specific task. In all cases, reference was made to the daily subject log in order to confirm the nature or possible cause of the outlier data. Although in most cases the procedure identified outliers on the poor performance side, it was also helpful in identifying certain subjects who changed their task performance strategies late in the study and performed much better than their typical performance.

The following narrative summarizes the magnitude of the outlier elimination for each task in addition to the two subjects completely eliminated from the database. The number of eliminated trials is from a total of 1362 training, baseline, and retest trials for the STRES tasks and 681 trials for each of the CTS and WRAIR PAB tasks and the STRES Reaction Time forms (a total of 17,706 trials). For the STRES battery, the summary is as follows: 23 trials involving 9 subjects for GRM, 7 trials involving 4 subjects for MTH, 4 trials involving 3 subjects for STN2, 5 trials involving 3 subjects for STN4, 5 trials involving 3 subjects for SPA, and 9 trials involving 4 subjects for TRK. In addition, all Unstable Tracking data for one subject were eliminated. Across all forms for STRES Reaction Time, a total of 16 trials involving 11 subjects were removed, in addition to all of one session for one subject and all of the CODED form data for one subject. Outlier removal for the CTS was 12 trials involving 7 subjects for GR, 3 trials involving 3 subjects for MP, 2 trials with 2 subjects for MS, 1 trial for SP, and 4 trials with 4 subjects for UT. As with the STRES, all UT data for one subject (the same subject) were eliminated. WRAIR PAB outlier screening consisted of 3 trials for the Time Wall task and 4 trials for Interval Production.

5.1.1 Normative Data from Baseline Trials

Table 13 presents baseline data summary statistics for the major dependent variables associated with each task. These data represent performance on the second day of baseline data collection (first trial of second day for STRES battery tasks). This particular trial was selected as being most representative of what the subjects could accomplish in terms of performance. The first STRES trial of the second day was selected in order to minimize the influence of fatigue which may have affected performance on the second trial. Means and standard deviations for the selected dependent measures associated with each STRES battery, CTS, and WRAIR PAB task are included. Comparable data (mean, standard deviation, median, lower quartile, and upper quartile) for each training, baseline, and retest trial are provided in separate appendices for each battery.

The data presented in Table 13 reveal a high degree of consistency between similar tasks and even across different tasks for some variables (e.g., response time for the STRES Reaction Time task and percentage correct measures in general). The apparent consistency between similar tasks across different batteries is encouraging and suggests that these variations of the same task may relate well to one another. The major exception to this general correspondence lies in the Unstable Tracking data where values for the CTS version are more than twice the values derived from the STRES version. This task presents a unique problem because the nature of its presentation and the calculation of the performance measures are highly dependent on the specific software algorithm used. Thus, it is difficult to infer comparability or the lack thereof across these task versions based on strict numerical data values. Data trends and general response characteristics serve as more important comparative indices along with the usual measures. Detailed comparisons of similar tasks across batteries are addressed in Section 5.3.

The high degree of similarity across the response time measures for the STRES Reaction Task suggests that these various forms of the task all require similar levels of response speed. The consistency with regard to percentage correct measures in general suggests that error rates for these tasks were fairly low across all tasks in all batteries. This uniformly high level of performance presents some difficulties with respect to reliability. These issues will be explored further in the reliability analysis Section 5.2.

5.1.2 STRES Battery

Discrete Response STRES Tasks. Figures 4 and 5 present the mean response time for correct responses (RT) and the percentage correct (PC) data respectively for the discrete response STRES battery tasks. These data represent group performance on each training day (Training 1a to 5b) and each baseline day (Baseline 1a to 2b). On each test day, subjects completed two trials for

each of the STRES tasks except the Reaction Time task. In the figures, the first trial is designated by the suffix 'a', and the second trial by the suffix 'b'.

Table 13. Means and Standard Deviations for Baseline Data (Baseline Day 2).

Battery	Task	Mean	Std. Dev.	Mean	Std. Dev.
STRES		Response Time		Percentage Correct	
	GRM	4507	1275	96%	7%
	MTH	1522	440	98%	4%
	STN2	476	68	98%	2%
	STN4	552	99	97%	3%
	SPA	947	231	95%	4%
	RCT	Response Time		Percentage Correct	
	1 - BASIC	562	93	98%	3%
	6 - BASIC	588	124	97%	4%
	2 - CODED	638	112	96%	4%
	3 - UNCERT	678	158	98%	5%
	4 - DOUBLE	595	104	97%	4%
	5 - INVERT	651	120	96%	4%
		Edge Violations		RMS Error	
	TRK	0.3	0.8	5.9	4.1
CTS		Response Time		Percentage Correct	
	GR	4855	1454	96%	5%
	MP	1703	502	96%	5%
	MS4	615	118	98%	2%
	SP	836	220	93%	4%
		Edge Violations		RMS Error	
	UT	1.1	2.6	11.7	7.0
WRAIR		Response Time		Percentage Correct	
	MAN	1179	466	97%	4%
		Interval Mean		Interval S.D.	
	INT	1024	129	72	33
	TIM	9670	868	411	256

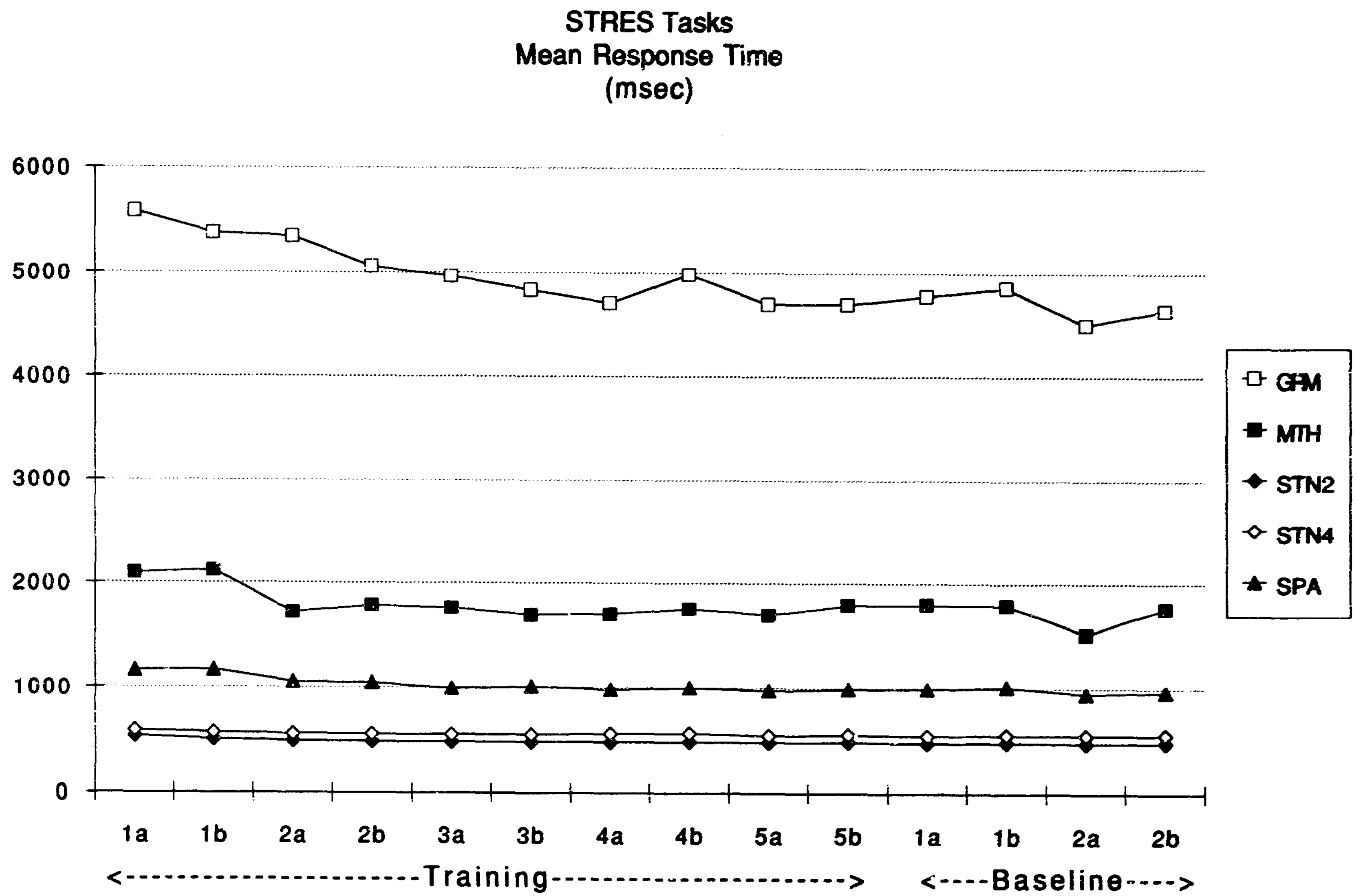


Figure 4. Mean Response Time for Discrete Response STRES Tasks.

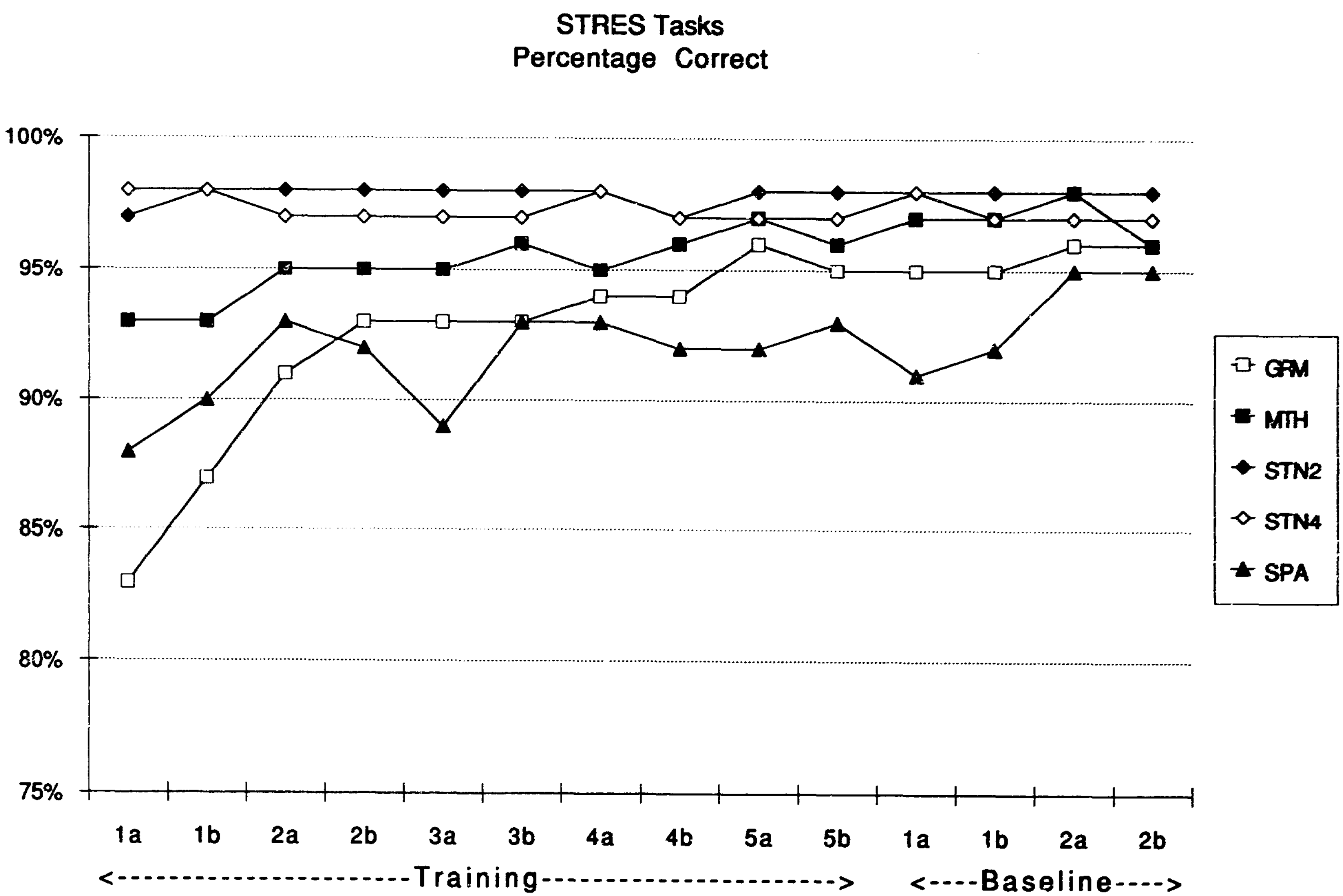


Figure 5. Mean Percentage Correct for Discrete Response STRES Tasks.

The figures suggest that stable performance was reached quite rapidly for the majority of tasks. In fact, many tasks appeared to demonstrate little improvement after the second day of training. The Grammatical Reasoning task was a notable exception. This task appeared to take longer to learn in terms of achieving both stable response time and percentage correct. Stable performance did appear to be reached by the fifth training day. The obvious explanation for this difference is revealed in the general separation between Grammatical Reasoning and the other tasks in terms of response time, as well as numerous anecdotal comments from subjects. That is, the Grammatical Reasoning task is generally viewed as the most challenging of all the tasks.

Reaction Time Task. Figures 6 and 7 present the mean response time for correct responses and percentage correct data for the STRES Reaction Time task. This task is somewhat unique. The subject provides responses to visual stimuli in a variety of forms. Each of these forms of the task is similar in that responses are limited to keyboard button presses with the index and second fingers of the right and left hands. However, the instructional sets are different for each form (see AGARD, 1989). In the Basic form, the subject uses the left hand if the stimulus appears on the left side of the screen and the right hand if the stimulus is on the right side. Within each hand, the leftmost finger is used if the stimulus is a '2' or '3' and the rightmost finger is used if the stimulus is a '4' or '5'. A Basic series is presented at the beginning (1 - BASIC) and at the end (6 - BASIC) of the sequence. In the second series (2 - CODED), the stimulus quality is degraded to one of four levels. Time uncertainty for the appearance of the stimulus is introduced in the third series (3 - UNCERT) by varying the interstimulus interval between 2 and 10 seconds. Double responses are required in the fourth series (4 - DOUBLE) such that the subject must press a sequence of three keys with the same hand beginning with the key for the correct response. In the fifth series (5 - INVERT), the screen-side to hand response mapping is inverted, i.e., stimuli on the left side of the screen require responses with the right hand and vice-versa.

The general similarity in the requirements of the various forms of the task probably accounts for the general uniformity in response time across the forms (see Table 13). Like many of the other STRES and CTS tasks, the percentage correct measures for the various forms of the STRES Reaction Time task were uniformly high (see Table 13). Again, fairly stable performance on both of these measures is seen beyond the second day for most tasks.

Although not statistically significant, it is interesting to note that during the last day of training and the two days of baseline testing the mean response times for the various forms of this task retained a consistent ordering. The first Basic series (1) provided the fastest RT followed by the final Basic series (6). Next in order were the Double key press (4), the Coded stimulus (2), and the Inverted hands (5) forms. Finally, the time Uncertain form (3) had the slowest RT.

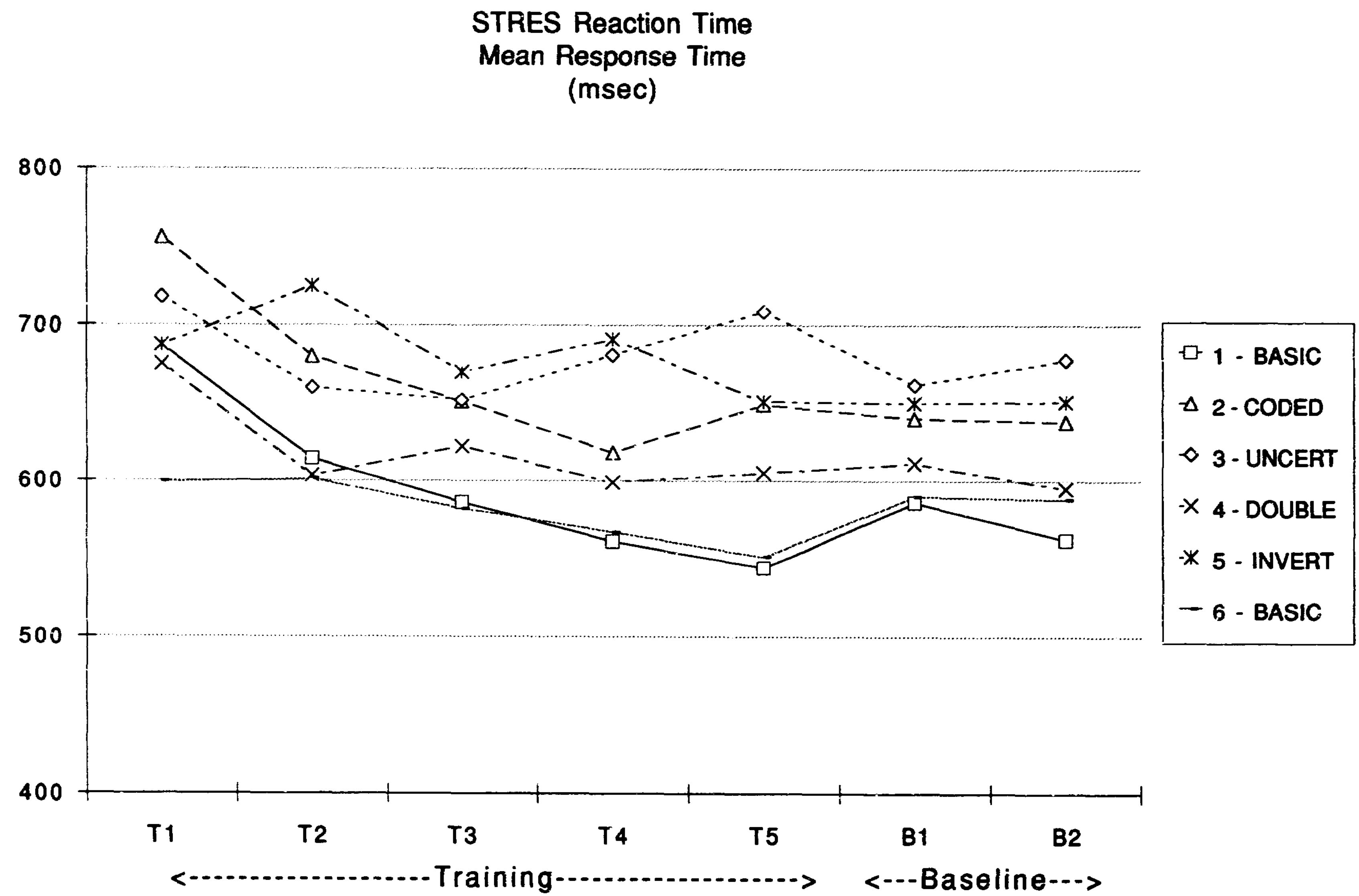


Figure 6. Mean Response Time for STRES Reaction Time Task.

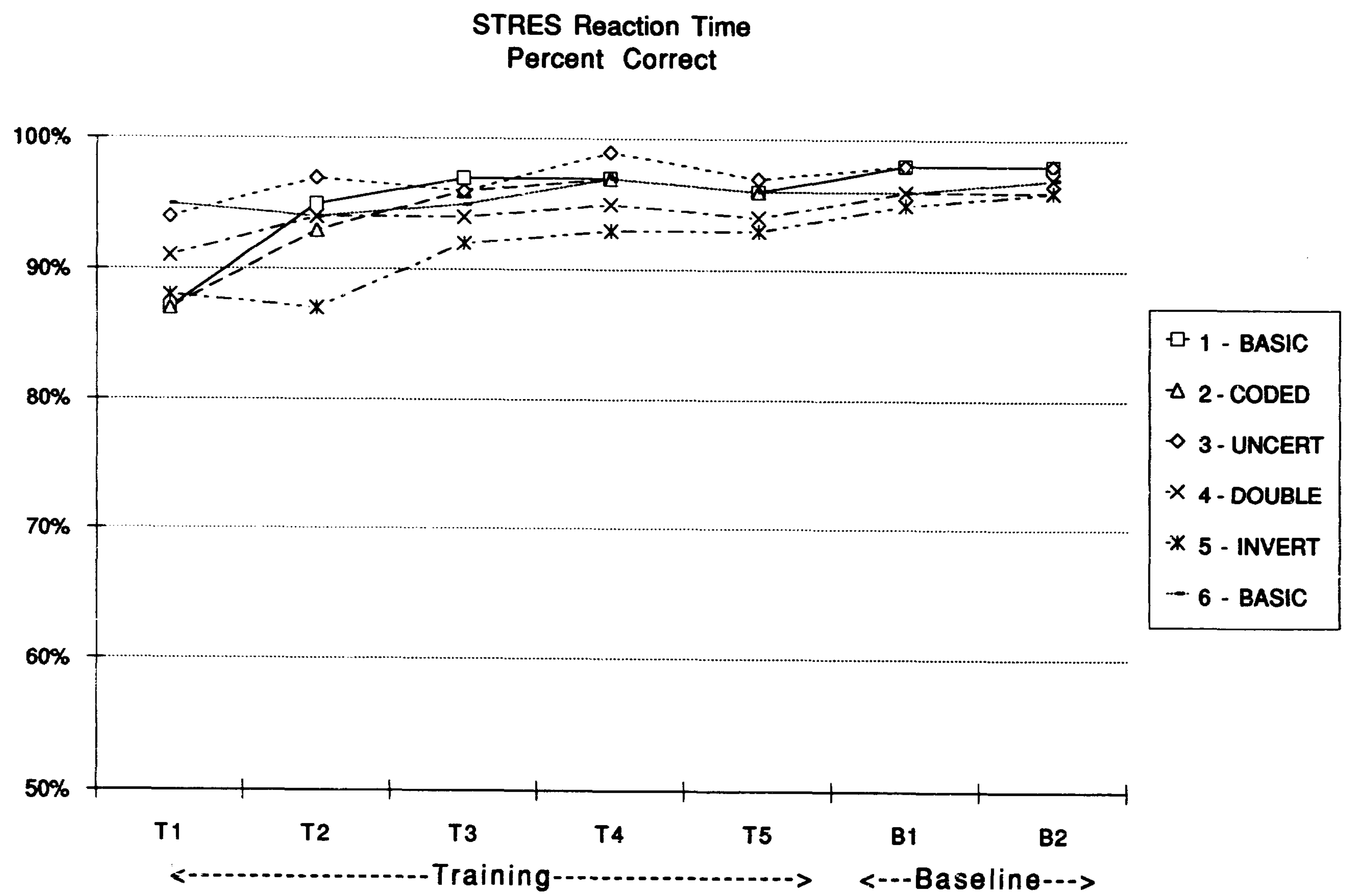


Figure 7. Mean Percentage Correct for STRES Reaction Time Task.

Unstable Tracking Task. Figure 8 presents the Edge Violation and RMS Error measures for the STRES battery Unstable Tracking task. Figure 8 suggests longer learning curves for the Unstable Tracking task, compared to other tasks. Stable performance in terms of Edge Violations appeared earlier (about the third training day) compared with RMS Error which does not appear to reach stability until the Baseline Testing days. This result is logical because the number of control losses at the periphery (Edge Violations) decreases early (and probably proportionally) as the subject, with continued practice, progressively reduces tracking variation about the center point (i.e., RMS Error). It is worth cautioning, however, that the Unstable Tracking task is among a small group of tasks that probably require more practice to attain stable performance than most other tasks assessed in this project. It is even more important to consider this issue when one recalls that, in this protocol, Unstable Tracking and Grammatical Reasoning were actually practiced three times per day (i.e., two STRES trials and one CTS trial per day).

Dual-Task Combination (COMBO). Those subjects who participated in the Response Deadline study performed a series of eleven trials of the COMBO task interspersed throughout the deadline testing. This task replaced the Unstable Tracking task in the subject's normal task sequence. Summary tables and graphs for the two component tasks are presented in Appendix C. A comparison with single-task baseline performance using Trial 1 from Baseline Day 2 indicates that subjects maintained the same level of performance on both tasks under dual-task conditions. The mean response time for the Memory Search task showed slight improvement over the course of the eleven trials.

Additional information regarding normative data for the STRES battery tasks is provided in Appendix C, which contains tabled values for training, baseline, and retest trials and individual graphs (mean, median, lower quartile, upper quartile) of training and baseline data for each task.

5.1.3 CTS Battery

Figures 9 and 10 present similar performance data for the CTS tasks that yield response time and percentage correct measures, respectively. These data, which represent group performance on each training day (T1-T5) and each baseline day (B1-B2), showed remarkable correspondence to the trends observed for similar STRES battery tasks. In this regard, the mean response time measures stabilized rapidly with the Grammatical Reasoning task lagging behind. Percentage correct measures were uniformly high across all tasks. This correspondence was expected because these are the tasks that are the most similar across the two batteries. As with the corresponding STRES tasks, these CTS task measures appear to reach reasonable levels of stability by the second or third day of training. The only exception appears to be the CTS Unstable Tracking task.

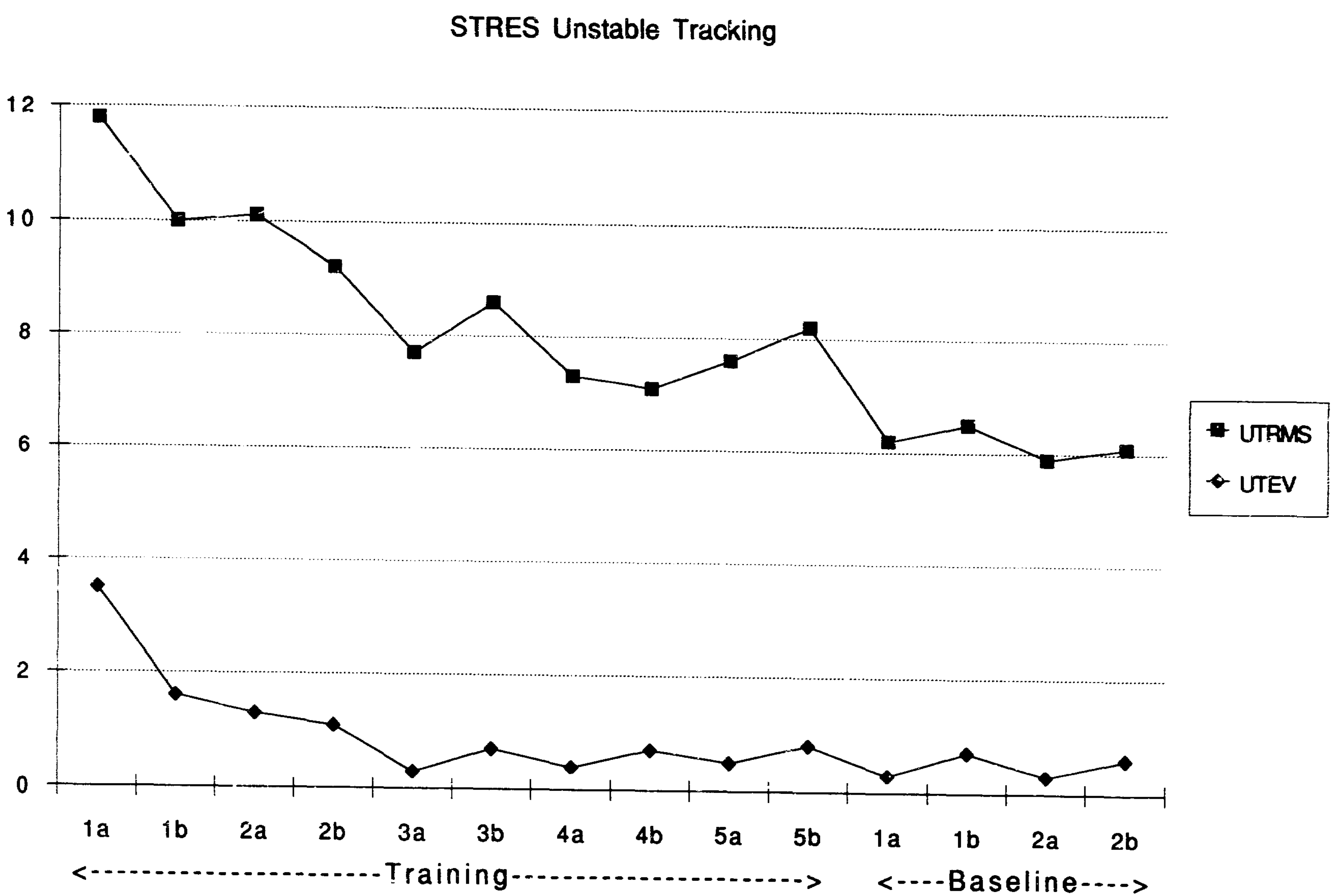


Figure 8. Mean RMS Error and Edge Violations for STRES Unstable Tracking Task.

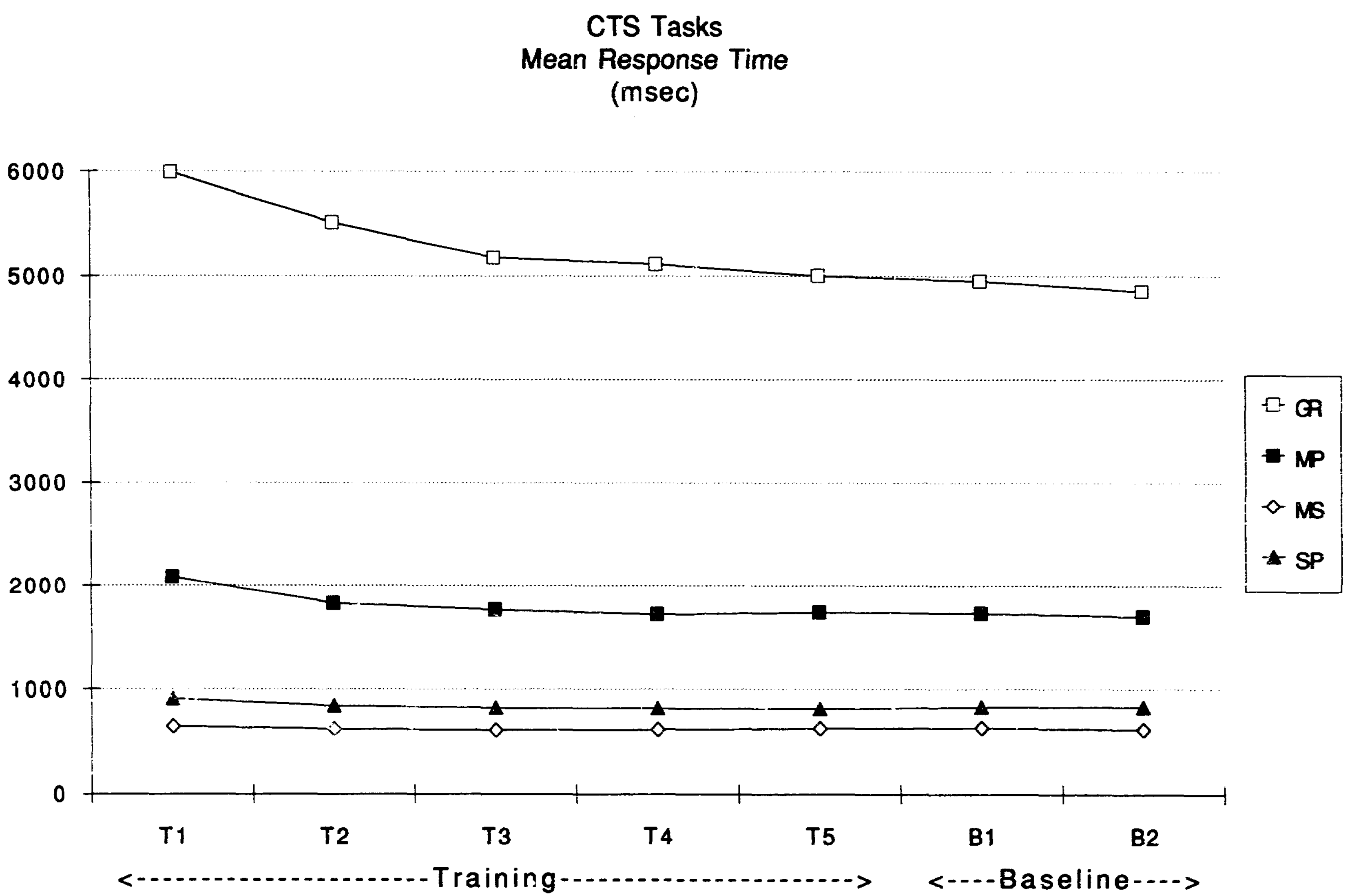


Figure 9. Mean Response Time for Discrete Response CTS Tasks.

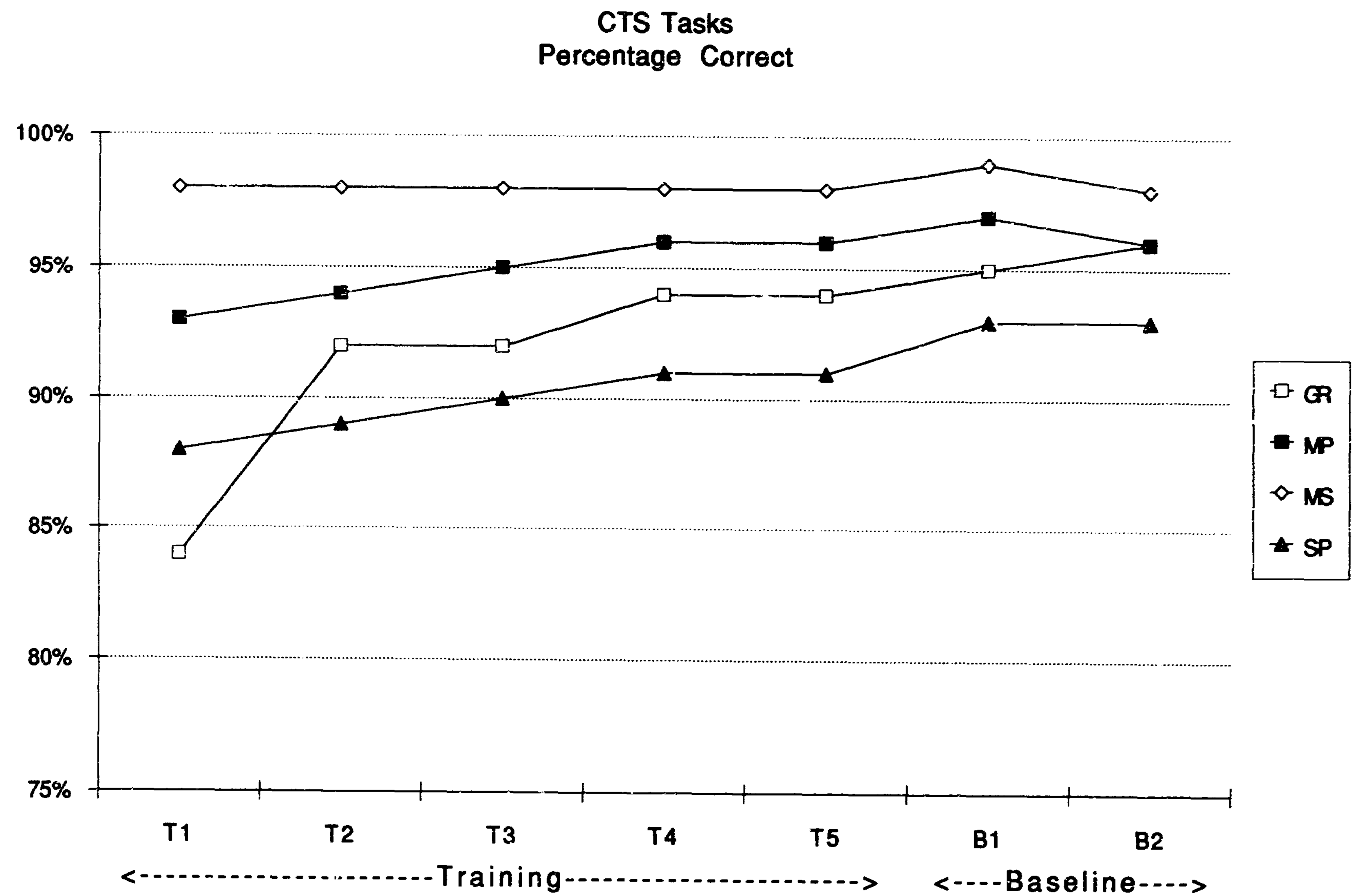


Figure 10. Mean Percentage Correct for Discrete Response CTS Tasks.

Data for the CTS Unstable Tracking task are presented in Figure 11. Unlike the data from the STRES Unstable Tracking task, the CTS data in Figure 11 represent only subjects from the University of Oklahoma. A major difference existed in the data obtained at the two testing sites. The discrepancy was traced to a difference in the manufacturer of the potentiometer in the tracking controllers used at Armstrong Laboratory and those used at the University of Oklahoma (even though both potentiometers meet the specifications). This was a regrettable, but entirely unpredictable, problem that resulted in the data from the Armstrong Laboratory being highly inconsistent with both the data collected at the University of Oklahoma, as well as previous CTS normative data collected at both locations. For that reason, only the CTS Unstable Tracking data from the University of Oklahoma were included in summaries and analyses.

In terms of response characteristics and trends, the CTS Unstable Tracking data were similar to the tracking data from the STRES battery. More training trials were needed to reach stability in tracking performance compared with other CTS tasks. Edge violations in the CTS data also appeared to stabilize before the RMS Error.

Additional information regarding normative data for the CTS tasks is provided in Appendix D, which contains tabled values for training, baseline, and retest trials and individual graphs (mean, median, lower quartile, upper quartile) of training and baseline data for each task.

5.1.4 WRAIR PAB

The mean response time and percentage correct measures for the WRAIR PAB Manikin task are presented in Figures 12 and 13, respectively. As with some of the other more difficult tasks noted previously, this task appears to require more training to achieve stable RT performance. It appears that additional increments in performance efficiency for percentage correct may not be important beyond trial four or five. However, this trend may be open to question for mean response time. Subjects appeared to show slight improvement in mean response time performance even through the baseline testing sessions.

Figures 14 and 15 present the means of the estimated time intervals and the means of the interval variability (standard deviation), respectively, for trials on the WRAIR PAB Time Wall task. Stable performance appeared to be attained quite rapidly for this task. Although there was a slight decrease in the mean time interval during the second through fifth training trials (compared with the first training trial), the mean intervals for baseline did not appear to be different from the intervals on the first training trial. Variability on this task showed improvement between the first and second trial, but was stable throughout the remainder of the training and baseline trials.

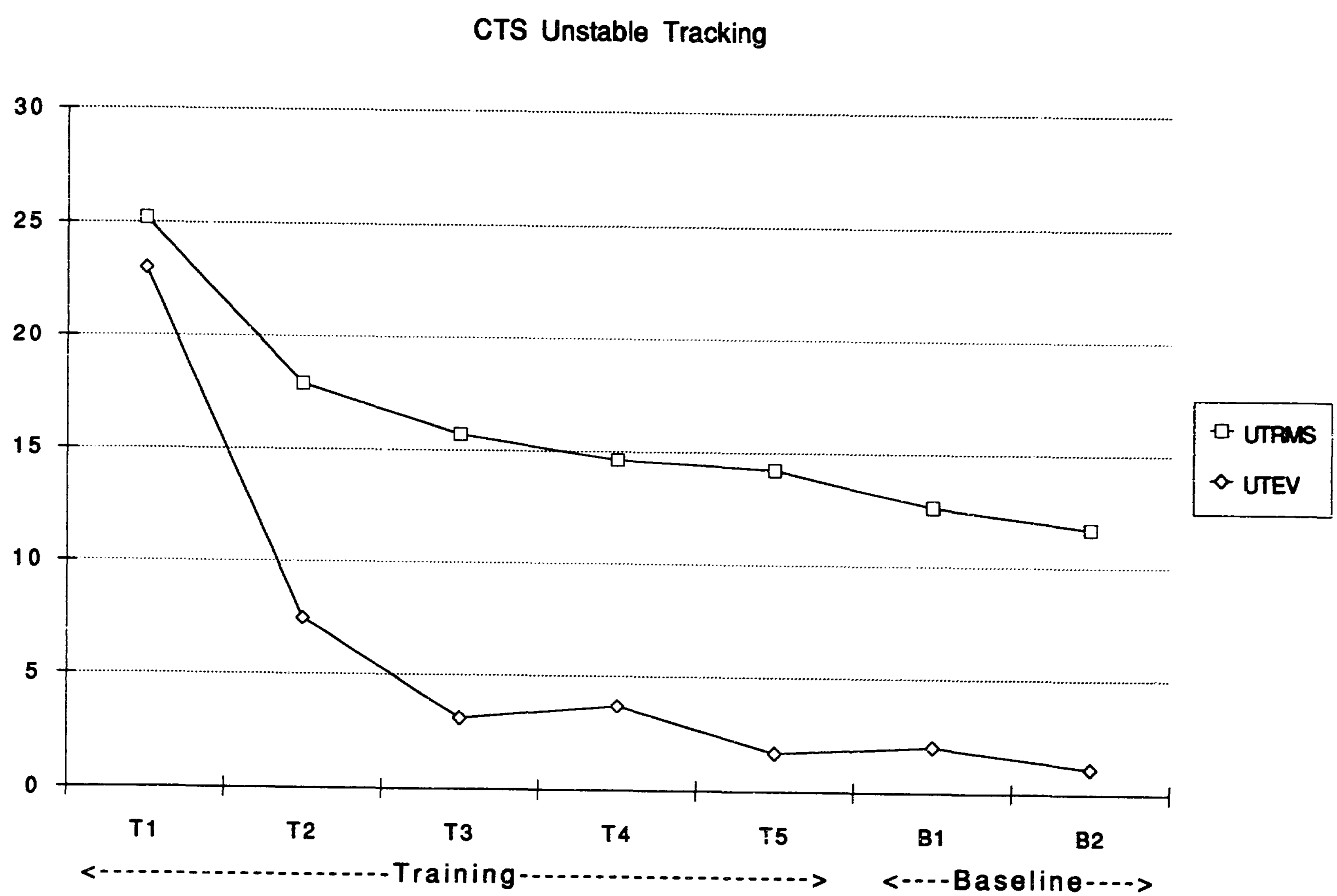


Figure 11. Mean RMS Error and Edge Violations for CTS Unstable Tracking Task.

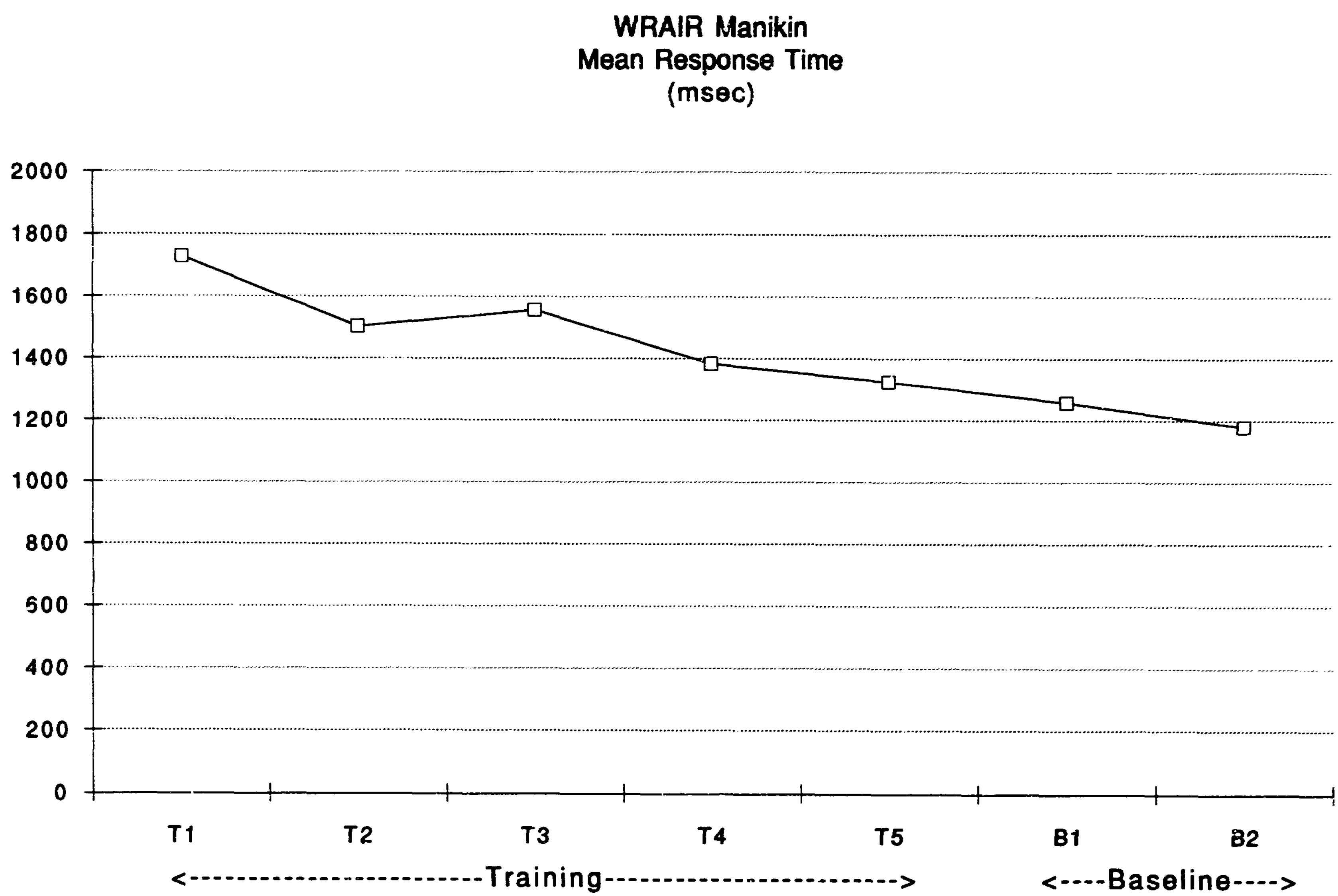


Figure 12. Mean Response Time for WRAIR PAB Manikin Task.

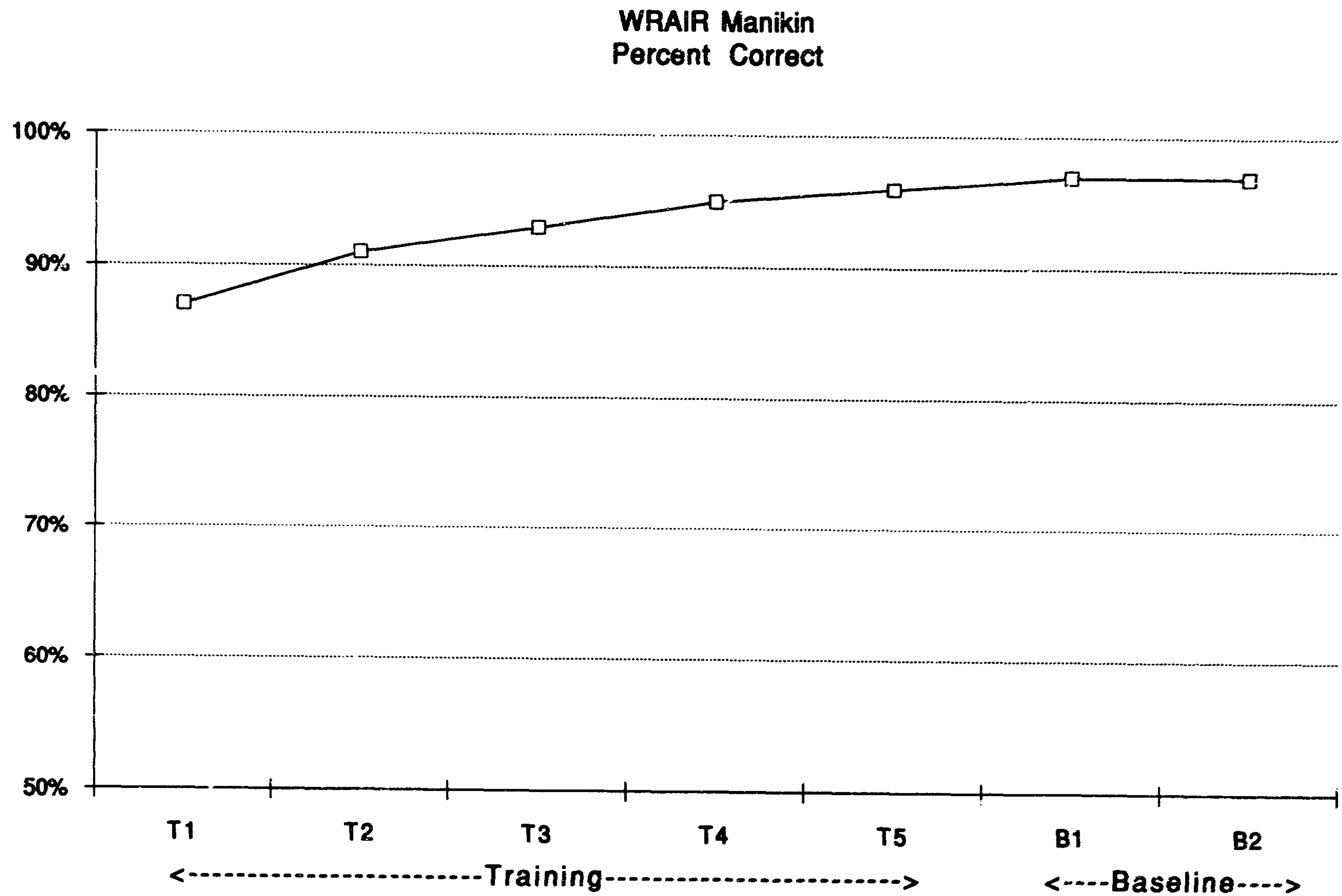


Figure 13. Mean Percentage Correct for WRAIR PAB Manikin Task.

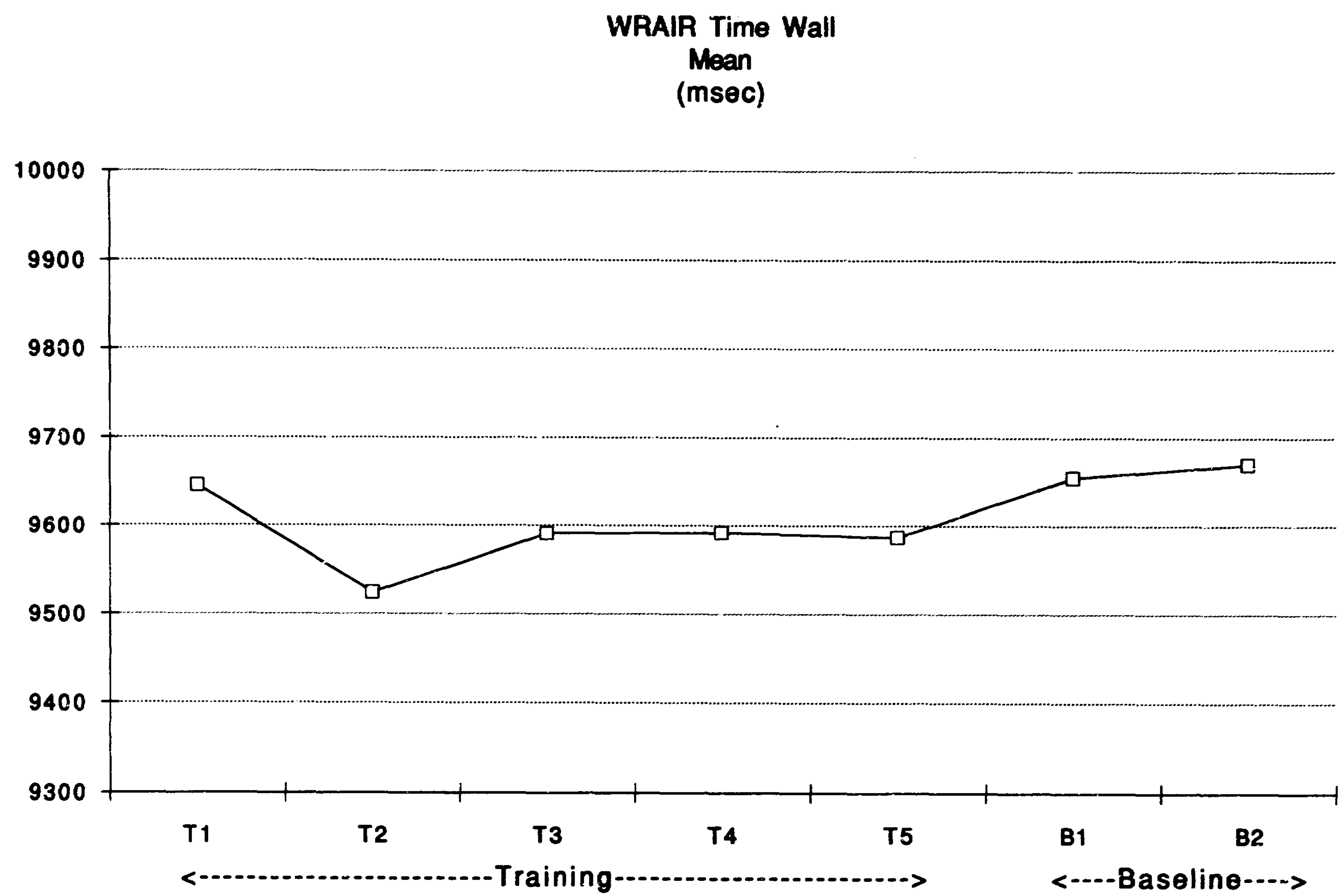


Figure 14. Mean Interval Length for WRAIR PAB Time Wall Task.

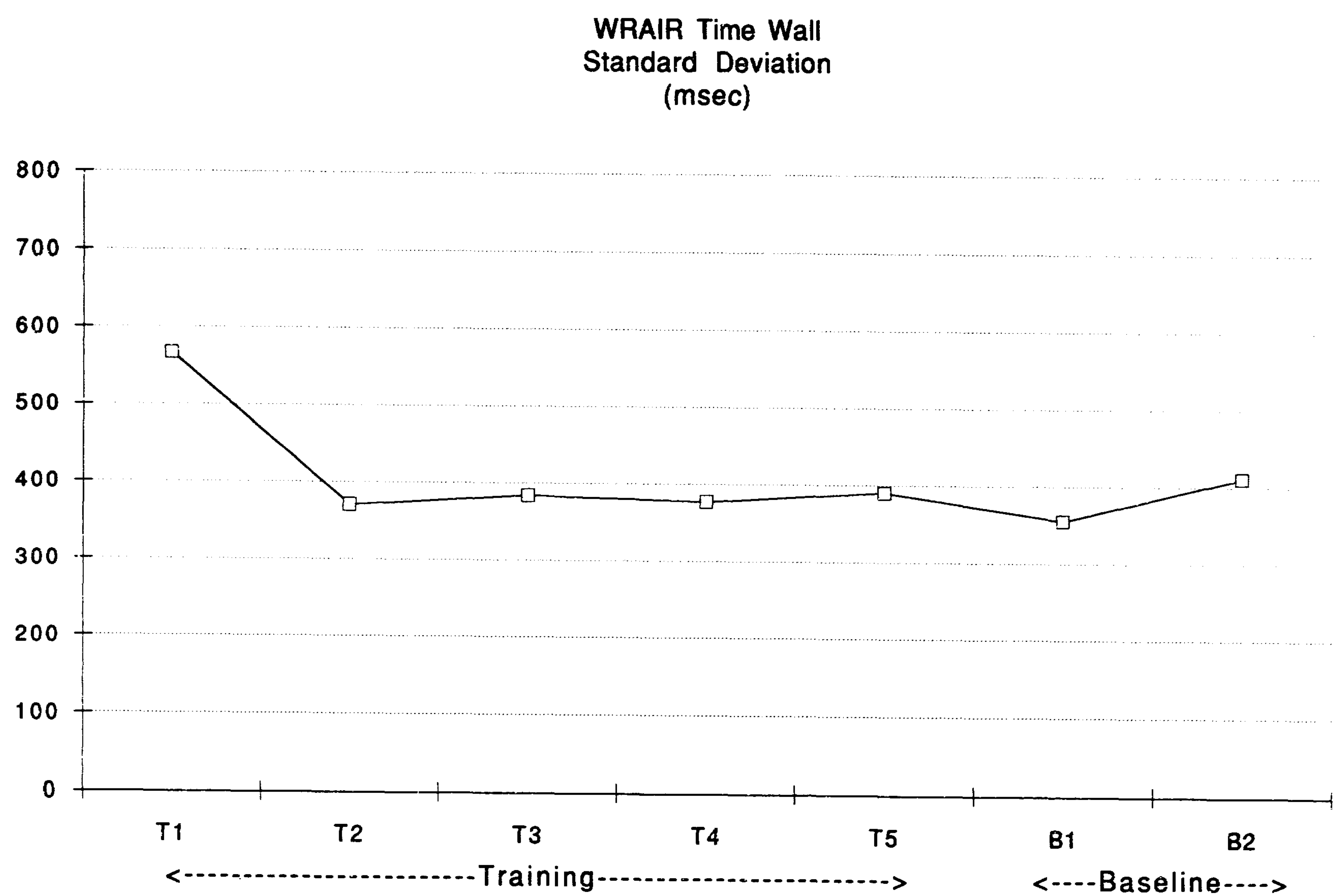


Figure 15. Mean Interval Variability for WRAIR PAB Time Wall Task.

The means of average interval length and interval variability (standard deviation) for the WRAIR PAB Interval Production task are presented in Figures 16 and 17. Like the Time Wall task, stable performance was reached rapidly on the Interval Production task. By possibly the third, and certainly the fourth trial, the mean interval measure appeared to reach asymptote. Variability appeared to stabilize within two trials.

Additional information regarding normative data for the WRAIR PAB tasks is provided in Appendix E, which contains tabled values for training, baseline, and retest trials and individual graphs (mean, median, lower quartile, upper quartile) of training and baseline data for each task.

5.1.5 Performance Percentile Groupings

Performance percentile groupings were calculated for each task within each battery. These performance percentile groupings provide estimates of the relevant dependent measures for performance categories ranging from Very Poor to Very Good in 20 percentile increments. These tables may be of particular interest to those researchers who wish to categorize their subjects (individuals or groups) based on the data from this study. Tables 14 and 15 present the performance percentile groupings for the STRES battery tasks. The performance percentile groupings for the CTS tasks are presented in Table 16, and the performance percentile groupings for the WRAIR tasks are presented in Table 17. A comparison of the current data for the STRES Reaction Time task with similarly presented data in AGARD (1989) reveals that the current sample of subjects had faster response times.

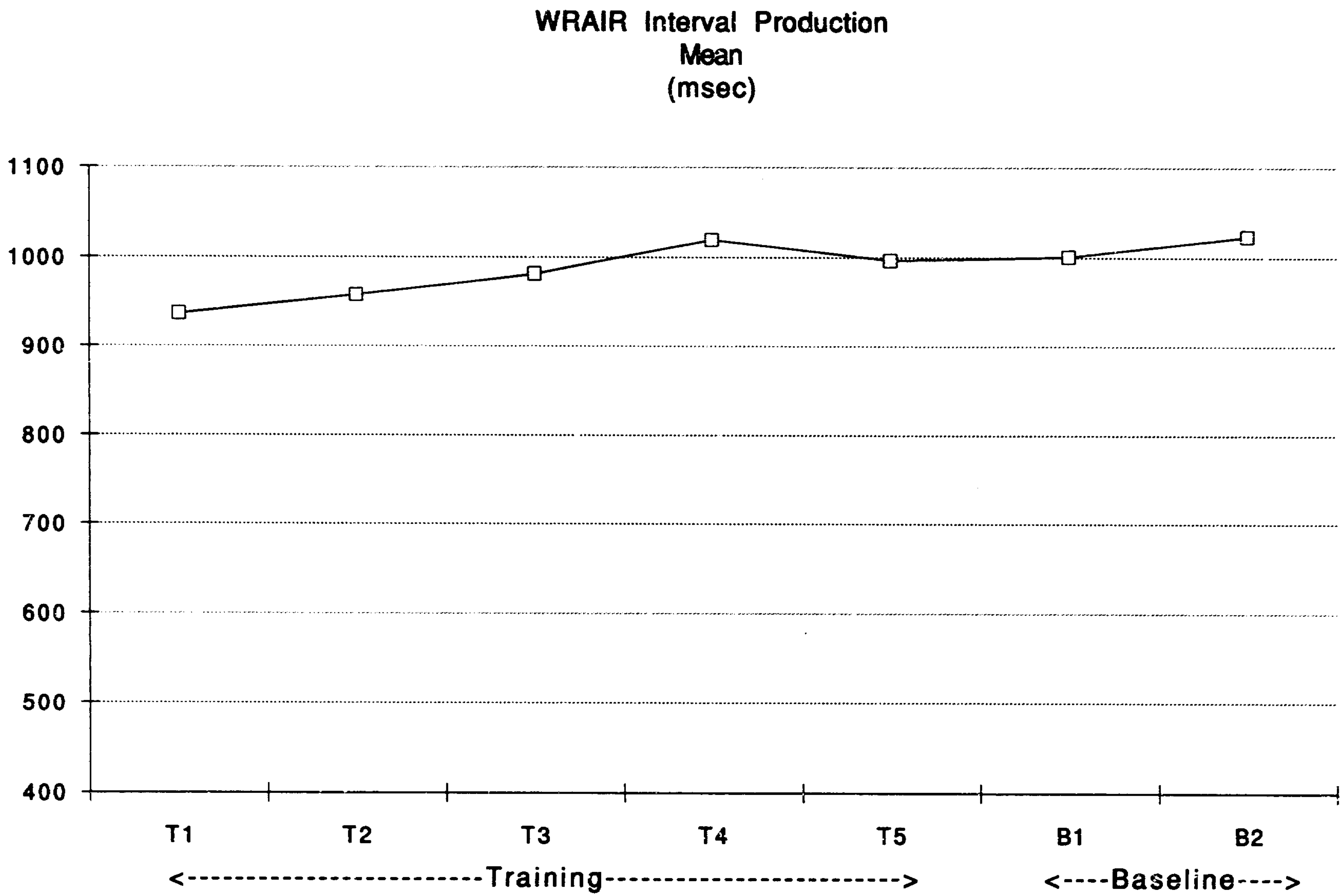


Figure 16. Mean Interval Length for WRAIR PAB Interval Production Task.

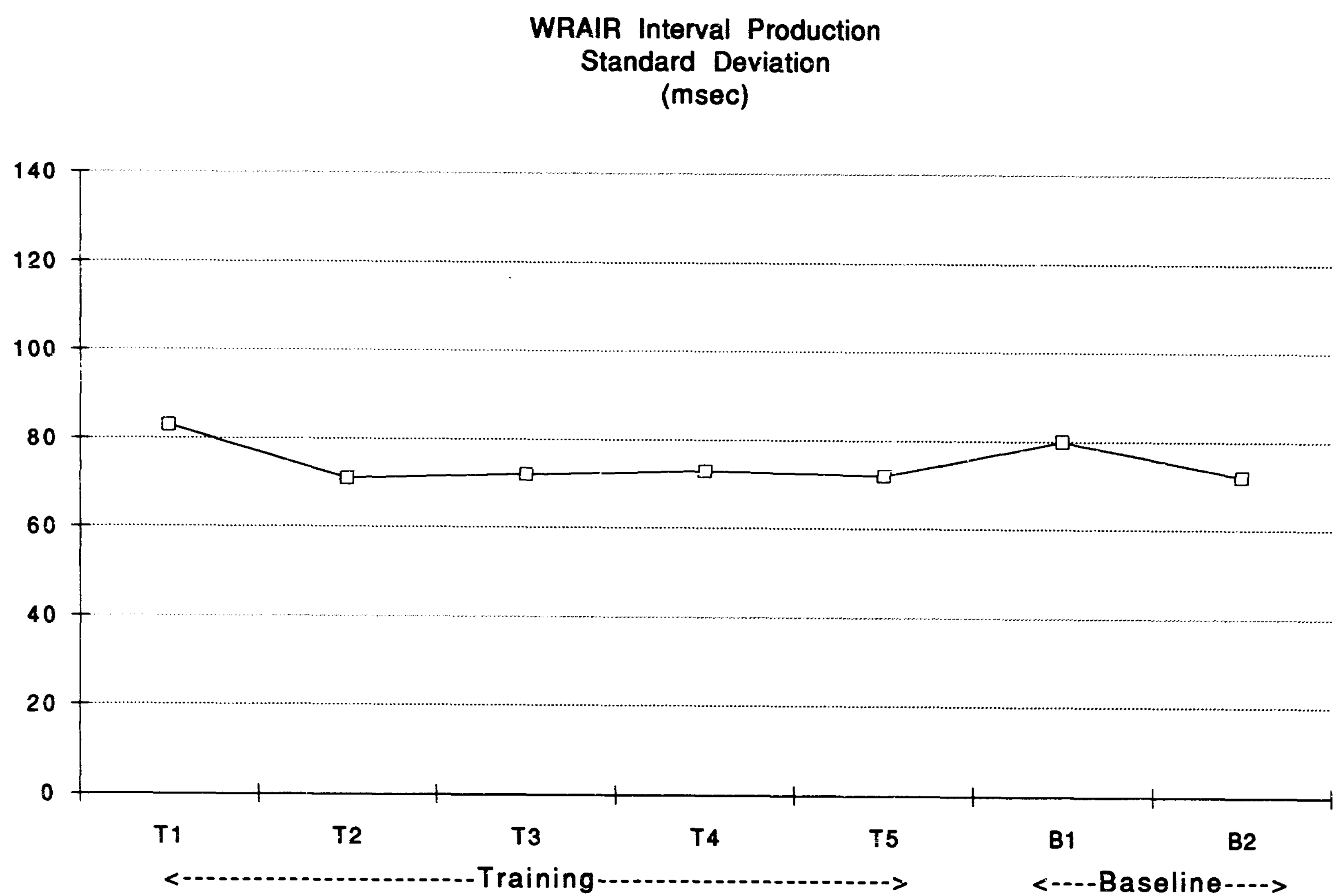


Figure 17. Mean Interval Variability for WRAIR PAB Interval Production Task.

Table 14. STRES Performance Percentile Groupings.

STRES Task Variable	Performance Category				
	Very Good 80%ile	Good 60-80%ile	Average 40-60%ile	Poor 20-40%ile	Very Poor 20%ile
GRM_Mean RT (msec)	<3304	3304-3943	3943-4697	4697-5731	>5731
GRM_Proportion Correct	1	.9773-1.000	.9706-.9773	.9333-.9706	<.9333
MTH_Mean RT (msec)	<1165	1165-1337	1337-1485	1485-1872	>1872
MTH_Proportion Correct	1	1	1	.9688-1.000	<.9688
STN2_Mean RT (msec)	<414	414-448	448-483	483-543	>543
STN2_Proportion Correct	1	.9912-1.000	.9815-.9912	.9652-.9815	<.9652
STN4_Mean RT (msec)	<463	463-514	514-563	563-629	>629
STN4_Proportion Correct	.9912-1.000	.9821-.9912	.9722-.9821	.9545-.9722	<.9545
SPA_Mean RT (msec)	<767	767-892	892-962	962-1083	>1083
SPA_Proportion Correct	1	.9722-1.000	.9444-.9722	.9143-.9444	<.9143
TRK_Edge Violations	0	0	0	0	>0
TRK_RMS Error	<2.6	2.6-4.2	4.2-5.5	5.5-8.8	>8.8

Table 15. STRES Reaction Time Task Performance Percentile Groupings.

REACT Block	Variable	Performance Category				
		Very Good 80%ile	Good 60-80%ile	Average 40-60%ile	Poor 20-40%ile	Very Poor 20%ile
1 - BASIC	Mean RT (msec)	<474	474-506	506-582	582-649	>649
	Proportion Correct	1	1	.9811-1.000	.9623-.9811	<.9623
6 - BASIC	Mean RT (msec)	<468	468-530	530-591	591-678	>678
	Proportion Correct	1	.9815-1.000	.9811-.9815	.9434-.9811	<.9434
2 - CODED	Mean RT (msec)	<549	549-573	573-646	646-729	>729
	Proportion Correct	1	.9811-1.000	.9623-.9811	.9434-.9623	<.9434
3 - UNCERT	Mean RT (msec)	<554	554-617	617-685	685-788	>788
	Proportion Correct	1	1	1	1	<1.000
4 - DOUBLE	Mean RT (msec)	<504	504-536	536-618	618-688	>688
	Proportion Correct	1	.9783-1.000	.9756-.9783	.9348-.9756	<.9348
5 - INVERT	Mean RT (msec)	<537	537-603	603-663	663-734	>734
	Proportion Correct	1	.9811-1.000	.9623-.9811	.9245-.9623	<.9245

Table 16. CTS Performance Percentile Groupings.

CTS Task Variable	Performance Category				
	Very Good 80%ile	Good 60-80%ile	Average 40-60%ile	Poor 20-40%ile	Very Poor 20%ile
GR_Mean RT (msec)	<3746	3746-4223	4223-4883	4883-6021	>6021
GR_Proportion Correct	1	.9730-1.000	.9592-.9730	.9375-.9592	<.9375
MP_Mean RT (msec)	<1263	1263-1503	1503-1723	1723-2105	>2105
MP_Proportion Correct	1	.9808-1.000	.9756-.9808	.9231-.9756	<.9231
MS_Mean RT (msec)	<503	503-580	580-629	629-709	>709
MS_Proportion Correct	1	1	.9787-1.000	.9762-.9787	<.9762
SP_Mean RT (msec)	<675	675-747	747-847	847-964	>964
SP_Proportion Correct	.9750-1.000	.9487-.9750	.9286-.9487	.9000-.9286	<.9000
UT_Edge Violations	0	0	0	0-3	>3
UT_RMS Error	<5.4	5.4-8.8	8.8-13.2	13.2-18.4	>18.4

Table 17. WRAIR PAB Performance Percentile Groupings.

WR Task Variable	Performance Category				
	Very Good 80%ile	Good 60-80%ile	Average 40-60%ile	Poor 20-40%ile	Very Poor 20%ile
MAN_Mean RT (msec)	<826	826-1002	1002-1158	1158-1464	>1464
MAN_Proportion Correct	1	1	.9792-1.000	.9375-.9792	<.9375
TIM_Standard Deviation	<219	219-313	313-424	424-543	>543
INT_Standard Deviation	<46	46-57	57-71	71-93	>93

5.2 Reliability of UTC-PAB Measures

The testing protocol for this UTC-PAB normative database project provided the opportunity to assess reliability at a variety of time intervals. As noted in Section 4.1, baseline testing for all subjects was conducted on the first two days of the second week. Approximately one half of the subjects (*N* = 33) returned one week later (five days) for two additional days of baseline retesting and the other half of the subjects (*N* = 31) returned three weeks (nineteen days) later for two additional days of baseline retesting. These various testing intervals provided numerous opportunities for assessing test-retest reliability. For example, comparing baseline data from Day 1 in Week 2 to baseline data from Day 2 in Week 2 provided an assessment of test-retest reliability over 24 hours. Because most of the STRES battery tasks were administered twice per day, comparison of the first and second trials on Day 2 of Week 2 provided an opportunity to assess retest reliability over approximately 30 minutes. An assessment of retest reliability over approximately one week was possible by comparing baseline measurements on Day 2 of Week 2 with baseline measures collected during the first retest session during the third week of the project. Finally, test-retest reliability over approximately three weeks was provided by comparing baseline testing on Day 2 of Week 2 with the first day of retesting during Week 5.

The reliability computations were supplemented by analyses of variance to identify performance differences across trials. These analyses are summarized in Section 5.2.4.

5.2.1 STRES Battery Reliability

Because the testing protocol allowed for the collection of two trials per day on many of the STRES tasks, retest reliability estimates for 30 minutes were calculated for these tasks in addition to the 24 hour, one week, and two week intervals reported for other tasks in this project. Table 18 presents Pearson product-moment correlations between the second baseline testing day (first trial) and baseline testing days at each of the four time intervals outlined above. Some general trends emerge from these data. First, a number of tasks appear to have considerable reliability with regard to their mean response time measure. These include Grammatical Reasoning, Mathematical Processing, and Spatial Processing. Reliability for the response time measures of these tasks fall well within the acceptable range for reliabilities, (i.e., 0.80 to 0.90). The reliabilities for the response time measures of the two versions of the STRES Memory Search (Sternberg) task fall in a somewhat marginal category. These reliabilities in the 0.70 to 0.80 range are marginal, yet encouraging, especially when considering these are performance task reliabilities. However, they are not as strong as would be desired, probably due to the relatively easy nature of the task such that the performance scores of most subjects are closely grouped.

Table 18. STRES Battery Test-Retest Correlations with Baseline Day 2.

STRES Task Measure	Retest Interval			
	30-minute	24-hour	1-week	3-week
GRM Response Time	0.91	0.90	0.91	0.86
GRM Percentage Correct	0.33	0.29	0.67	0.20
MTH Response Time	0.86	0.88	0.84	0.83
MTH Percentage Correct	0.47	0.69	0.69	0.48
STN2 Response Time	0.80	0.75	0.78	0.71
STN2 Percentage Correct	0.58	0.59	0.58	0.48
STN4 Response Time	0.79	0.70	0.82	0.72
STN4 Percentage Correct	0.66	0.50	0.42	0.10
SPA Response Time	0.90	0.84	0.83	0.84
SPA Percentage Correct	0.32	0.46	0.41	0.04
TRK Edge Violations	0.11	0.11	0.26	0.04
TRK RMS Error	0.69	0.67	0.74	0.43

In general, the reliabilities for the percentage correct measures for all of the above mentioned tasks fall in an unacceptable category. This problem of uniformly low reliabilities is most likely explained by the fact that the percentage correct measure is subject to an extreme ceiling effect for most tasks. Due to the ceiling effect, so little variability is retained in this measure that reliability clearly becomes compromised.

The reliability figures for the STRES Unstable Tracking task are generally quite low. The reliability figures for Edge Violations are generally in the unacceptable category and can be explained by the fact that very few Edge Violations occur beyond the first few training trials. Again, a lack of variability, in this case a floor effect, precludes effective reliability measurement. The RMS Error value, while retaining more variability, does not provide impressive reliability indices. The reliability of this measure is moderate for fairly recent periods up to one week, but appears to drop off considerably at the three-week testing interval.

It should also be noted that for many of these measures, reliability remains fairly constant across all testing intervals. Usually, this suggests that acceptably stable performance can be expected for task performance up to a three-week testing interval.

Table 19. STRES Reaction Time Task Test-Retest Correlations.

Block	Task Measure	Retest Interval		
		24-hour	1-week	3-week
1 - BASIC	Response Time	0.86	0.82	0.89
	Percentage Correct	0.51	0.59	0.47
6 - BASIC	Response Time	0.84	0.76	0.86
	Percentage Correct	0.60	0.61	0.12
2 - CODED	Response Time	0.88	0.87	0.91
	Percentage Correct	0.70	0.18	0.80
3 - UNCERT	Response Time	0.78	0.63	0.73
	Percentage Correct	0.12	0.00	-0.06
4 - DOUBLE	Response Time	0.93	0.87	0.89
	Percentage Correct	0.50	0.18	0.61
5 - INVERT	Response Time	0.87	0.91	0.83
	Percentage Correct	0.50	0.58	0.45

Table 19 presents the reliability indices for the various forms of the STRES Reaction Time task. Because the STRES Reaction Time task was completed only once per day, the calculation of a thirty-minute test-retest reliability was not possible. Again, quite acceptable reliability indices were obtained for the response time measure for each of the STRES Reaction Time task forms. The one possible exception was the time uncertainty form (3 - UNCERT) which provided somewhat marginal reliability indices. These reliabilities appeared to remain quite stable across the 24-hour, one-week, and three-week testing intervals.

Reliability indices for the STRES Reaction Time task percentage correct measures were uniformly poor. There was considerable deviation in these indices, with some providing reliability estimates as high as 0.70 to 0.80 and others as low as near zero. Again, ceiling effects in the percentage correct measure preclude any adequate measure of reliability.

5.2.2 CTS Reliability

Table 20 presents the reliability indices for the various CTS tasks. The response time measure for Grammatical Reasoning, Mathematical Processing, Memory Search, and Spatial Processing

provided a high degree of reliability across all three testing intervals. The reliability figures for the percentage correct measures for these tasks were again quite low for the same reasons mentioned previously in the STRES tasks analysis.

Table 20. CTS Test-Retest Correlations with Baseline Day 2.

CTS Task Measure		Retest Interval		
		24-hour	1-week	3-week
GR	Response Time	0.90	0.90	0.93
GR	Percentage Correct	0.44	0.33	0.19
MP	Response Time	0.86	0.88	0.77
MP	Percentage Correct	0.56	0.41	0.41
MS	Response Time	0.81	0.80	0.88
MS	Percentage Correct	0.32	0.41	0.34
SP	Response Time	0.86	0.84	0.85
SP	Percentage Correct	0.28	0.33	0.17
UT	Edge Violations	0.77	0.10	0.32
UT	RMS Error	0.85	0.70	0.72

The 24-hour test-retest reliability for the RMS Error measure of the CTS Unstable Tracking task fell in the acceptable range. Subsequent reliabilities for one- and three-week intervals were marginally acceptable. The 24-hour reliability index for Edge Violations was marginally acceptable. However, the reliabilities for subsequent time intervals were unacceptably low.

5.2.3 WRAIR PAB Reliability

The reliabilities for the WRAIR PAB Tasks are presented in Table 21. The correlations for the mean response time for correct items in the WRAIR PAB Manikin task suggest a high degree of reliability across the various retest intervals. The reliabilities for the percentage correct measure are unacceptably low, undoubtedly due again to ceiling effects.

Reliability figures for the mean time estimation for the WRAIR PAB Time Wall task are generally quite acceptable although the three-week reliability is somewhat low. Reliabilities for the standard deviation measure are generally quite low. However, this measure is generally of less utility.

Table 21. WRAIR PAB Test-Retest Correlations with Baseline Day 2.

Task	Task Measure	Retest Interval		
		24-hour	1-week	3-week
Manikin	Response Time	0.89	0.89	0.95
	Percentage Correct	0.37	0.58	0.03
Time Wall	Interval Mean	0.87	0.94	0.79
	Interval Std. Dev.	0.29	0.68	0.44
Interval Prod	Interval Mean	0.58	0.57	0.57
	Interval Std. Dev.	0.40	0.34	0.43

Reliabilities for both mean interval and standard deviation for the WRAIR PAB Interval Production task are generally quite low. This probably is not surprising given the very small interval of time that the subjects are asked to produce. In the WRAIR PAB Time Wall task the subject is basically being asked to estimate over a 10-second time interval. In the WRAIR PAB Interval Production task, the subject is being asked to estimate one-second intervals. At one level, these data suggest that time estimation may decrease in reliability as the standard time interval being estimated decreases. Alternatively, the visual component of the Time Wall task may provide sufficient cueing to distinguish this from being strictly a time estimation task. Some subjects may consistently under-estimate the time while others consistently over-estimate the time, thus leading to a higher reliability.

5.2.4 Reliability Summary

The reliabilities for tasks with response time measures were generally acceptable and in many cases fairly impressive. These reliabilities were sustained across time intervals from 30 minutes to three weeks. Due to ceiling effects, accuracy measures such as percentage correct appeared to have very poor reliability. However, the percentage correct measures could increase in reliability if the testing conditions were changed to increase the difficulty of the task either through environmental or task manipulations. Of the various tasks evaluated in this project, the WRAIR PAB Interval Production Task provided the least evidence of reliability. No reliability index of any measure in any testing interval exceeded 0.58 for this task. Also, of some concern are the measures of tracking ability from both the STRES and CTS batteries. Edge violations are generally unreliable in a traditional sense due to the floor effect and RMS Error values are reasonably reliable on the CTS Unstable Tracking task, but only marginally so on the STRES battery version of this task.

5.3 Statistical Analysis of Trial Differences

Analysis of variance was used to determine the statistical significance of any performance differences that existed across trials for three situations. The first situation was to verify the existence of a training effect. The second series of analyses identified differences across the two (CTS, STRES RCT, WRAIR PAB) or four (STRES) Baseline trials. The third series examined differences in performance between Baseline trials and one-week or three-week retest trials. With respect to the third series, this type of analysis helps to determine if performance has remained stable even when the test-retest correlations are low for other reasons. In many ways, this approach is more appropriate when measures exhibit either ceiling or floor effects and is also appropriate in other situations.

5.3.1 Training Trials

With few exceptions, there were significant improvements in performance across training trials, with the greatest changes occurring between Trial 1 and Trial 2. For this reason, Trial was included as a within-subjects factor in all ANOVA's used to investigate the various issues of concern (i.e., test administration, battery sequence, etc.). The only task measures which did not demonstrate a statistically significant training effect were (1) STRES Memory Search percentage correct at both the two-character ($p = 0.67$) and four-character ($p = 0.08$) levels, (2) CTS Memory Search percentage correct ($p = 0.94$), (3) WRAIR PAB Time Wall mean interval ($p = 0.23$), and (4) WRAIR PAB Interval Production standard deviation ($p = 0.29$). These results would be expected in the case of the Memory Search task on which subjects provide a high level of accuracy starting with the very first trial, and demonstrate performance improvement on the task by providing faster responses with no change in accuracy. The WRAIR PAB Time Wall and Interval Production tasks in some sense measure an inherent skill that does not improve with practice.

For most tasks in this project, performance appeared to reach asymptotic levels by approximately the third day of practice. However, it should be noted that this observation is based largely on those tasks represented in both the STRES and CTS batteries, between which some transfer of skill may have occurred during training. Thus, learned in isolation, each of these tasks may require more practice to attain stability. Estimates from an earlier CTS normative study suggest four to five trials of practice is the minimum needed to obtain reasonably stable performance (see Schlegel and Gilliland, 1990). More practice is needed on what appears to be the more difficult tasks (i.e., Grammatical Reasoning, Unstable Tracking, etc.). These tasks appeared to require at least five trials of practice.

5.3.2 Baseline Trials

A summary of the few statistically significant baseline trial effects is presented in Table 22. Of the forty measures analyzed, only nine demonstrated some statistically significant variation across the multiple baseline sessions. Keep in mind that a collection of univariate ANOVA's was performed with no project-wise control of the Type I error level. Thus, one would expect some significant results by chance. In general, for those measures where a difference was evident, Baseline Day 2 exhibited better performance than Baseline Day 1. On the STRES tasks, the first trial of Baseline Day 2 exhibited better performance. A similar though not statistically significant trend existed with other tasks, indicating the small but continuing effect of learning.

Table 22. Summary of Significant Baseline Trial Differences.

Task Measure	p > F	Tukey Test (.01)* (improved performance -->)
STRES-GRM Mean RT	0.0001	B1b B1a B2b B2a
STRES-MTH Mean RT	0.0001	B1a B1b B2b B2a
STRES-MTH Proportion Correct	0.0498	B2b B1a B1b B2a
STRES-SPA Mean RT	0.0061	B1b B1a B2b B2a
STRES-SPA Proportion Correct	0.0001	B1a B1b B2b B2a
STRES-RCT BASIC Block 1 RT	0.0009	B1 B2
STRES-RCT DOUBLE Block 4 RT	0.0022	B2 B1
WRAIR-MAN Mean RT	0.0074	B1 B2
WRAIR-TIM Standard Deviation	0.0166	B2 B1

5.3.3 Retest Trials

Figures 18 and 19 present typical results from the one-week and three-week retest trials. An initial statistical comparison of the performance of the two subject groups using only the training and baseline data confirmed that there were no significant differences between the groups with respect to any of the task measures. Having confirmed the comparability of the two subject groups, separate analyses of variance were conducted to compare baseline and retest performance for the one-week retest group and the three-week retest group. The number of statistically significant differences were few and often revealed greater variability among the baseline trials

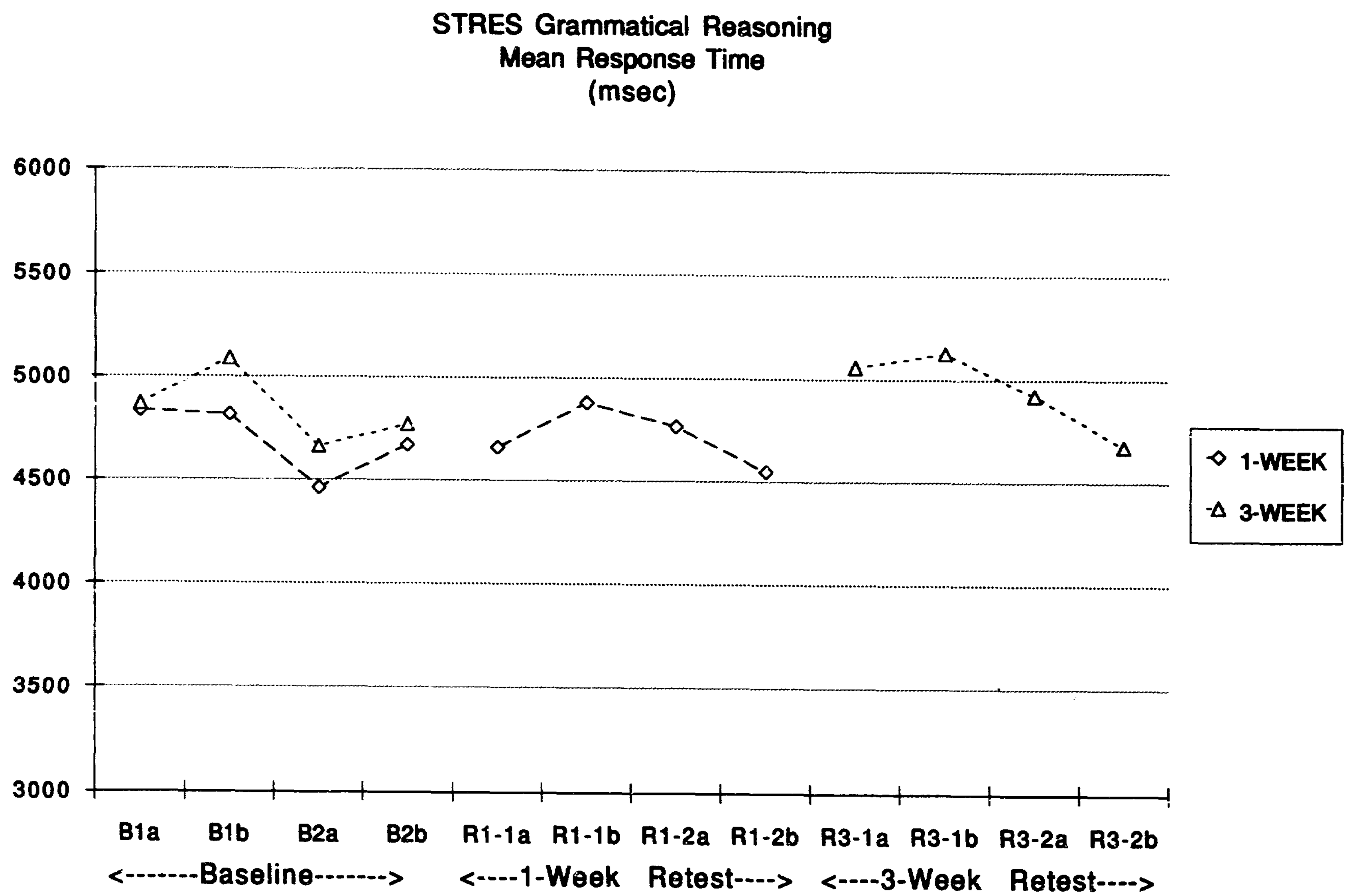


Figure 18. One-Week and Three-Week Retest Data for STRES Grammatical Reasoning Response Time.

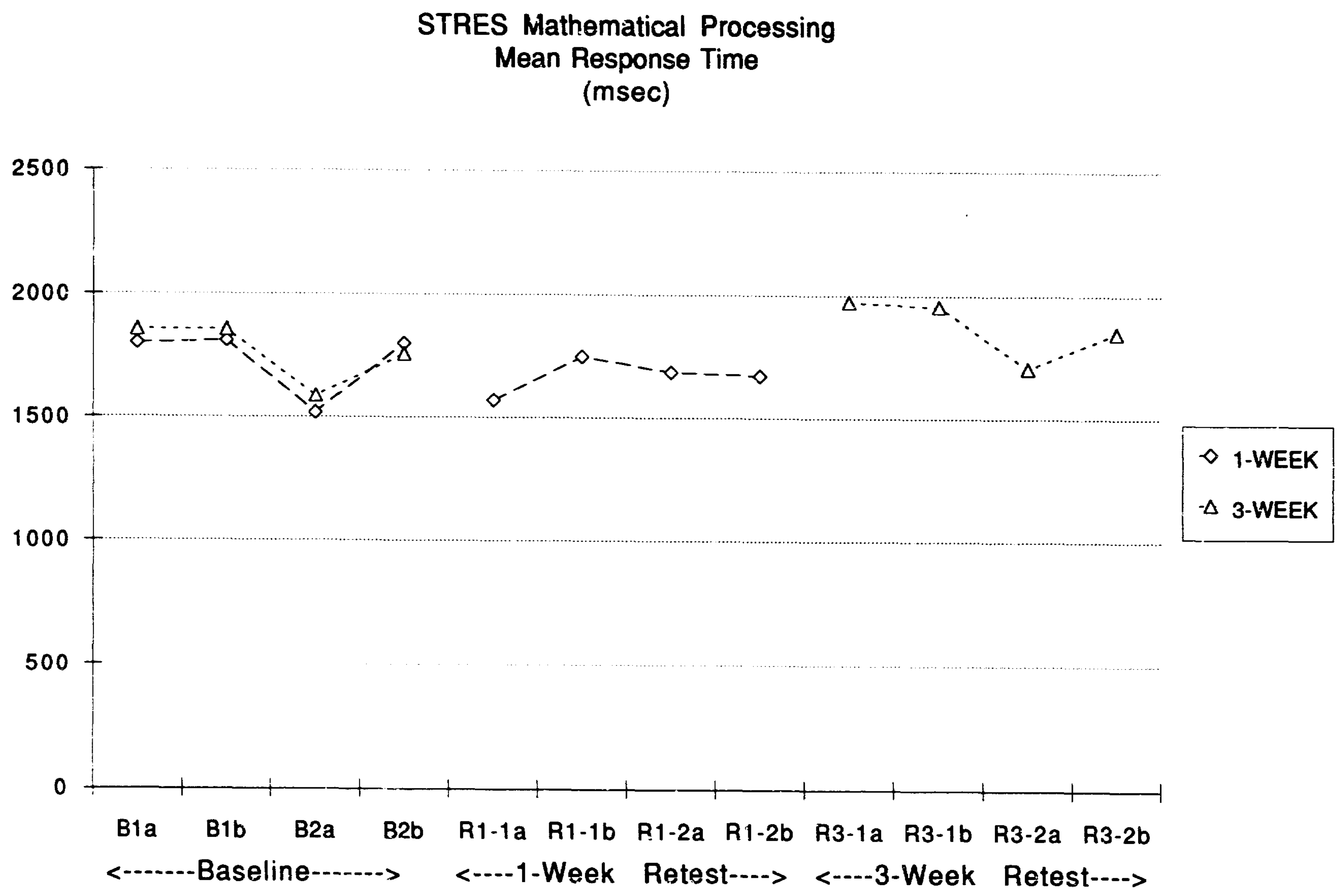


Figure 19. One-Week and Three-Week Retest Data for STRES Mathematical Processing Response Time.

themselves than between the baseline and retest trials. In general, when differences occurred, the first retest trial demonstrated slightly poorer performance than baseline but the second retest trial indicated better performance than baseline, indicating that subjects recovered quickly and perhaps continued to improve with only one additional practice trial. Graphs comparing baseline, one-week, and three-week performance for all tasks are provided in Appendix F.

5.4 Comparison of Similar Tasks Across Batteries

Due to the fact that each daily test period consisted of one session on the CTS tasks but two sessions on the comparable STRES tasks, caution must be exercised in making an appropriate comparison of the two batteries. As mentioned previously, it is unclear to what extent training on the CTS transferred to the STRES and vice-versa. An approximately equal number of subjects performed the batteries in the order CTS-STRES (37) and STRES-CTS (42) in an attempt to balance the transfer effect.

Although one may argue differently, the authors believe that the most appropriate approach is to compare CTS and STRES performance on a block-by-block basis regardless of the day on which the data were obtained. This would allow the seven blocks of CTS data (five training days plus two baseline days) to be compared with the first seven blocks of STRES data (first three and a half days of training). Comparisons of this nature are provided for the four discrete response tasks (GR/GRM, MP/MTH, MS/STN4, and SP/SPA) in Figure 20 (Mean Response Time) and Figure 21 (Percentage Correct). The comparison for the tracking tasks (UT/TRK) is presented in Figure 22. Individual graphs for each task using expanded scales are provided in Appendix G.

It is clear that there exists very good correspondence between the CTS and STRES implementations of most tasks. Where minor differences exist, accompanying distinctions in task implementation can be readily found to explain the differences. A discussion of each task follows.

For the Grammatical Reasoning task, there was no statistically significant difference between the batteries for the RT measure. CTS GR had a slightly higher initial RT, but this difference did not persist beyond Block 1. There was a statistically significant difference in Percentage Correct ($F_{1,78} = 11.99, p = 0.0009$) with the CTS GR yielding slightly better performance. One notable difference in implementations is the use of the words *PRECEDES* and *FOLLOWS* in CTS vs. *BEFORE* and *AFTER* in STRES. This difference may have caught subjects off guard on the first CTS trial. For the Mathematical Processing task, there were absolutely no performance differences for either RT or PC. The STRES and CTS implementations of these tasks are essentially identical with only minor differences of screen character size and response device.

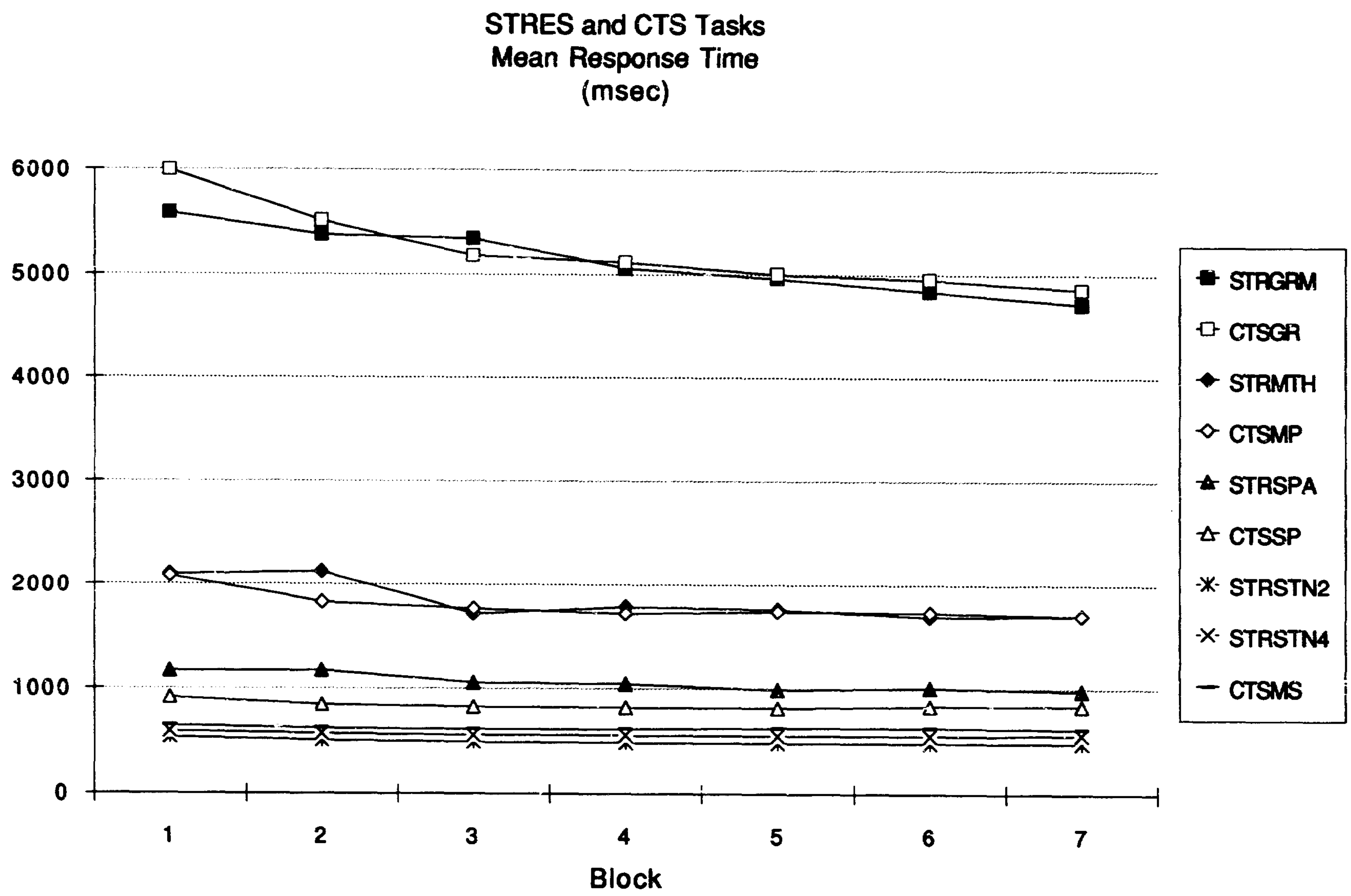


Figure 20. Mean Response Time Comparison of STRES and CTS Discrete Response Tasks.

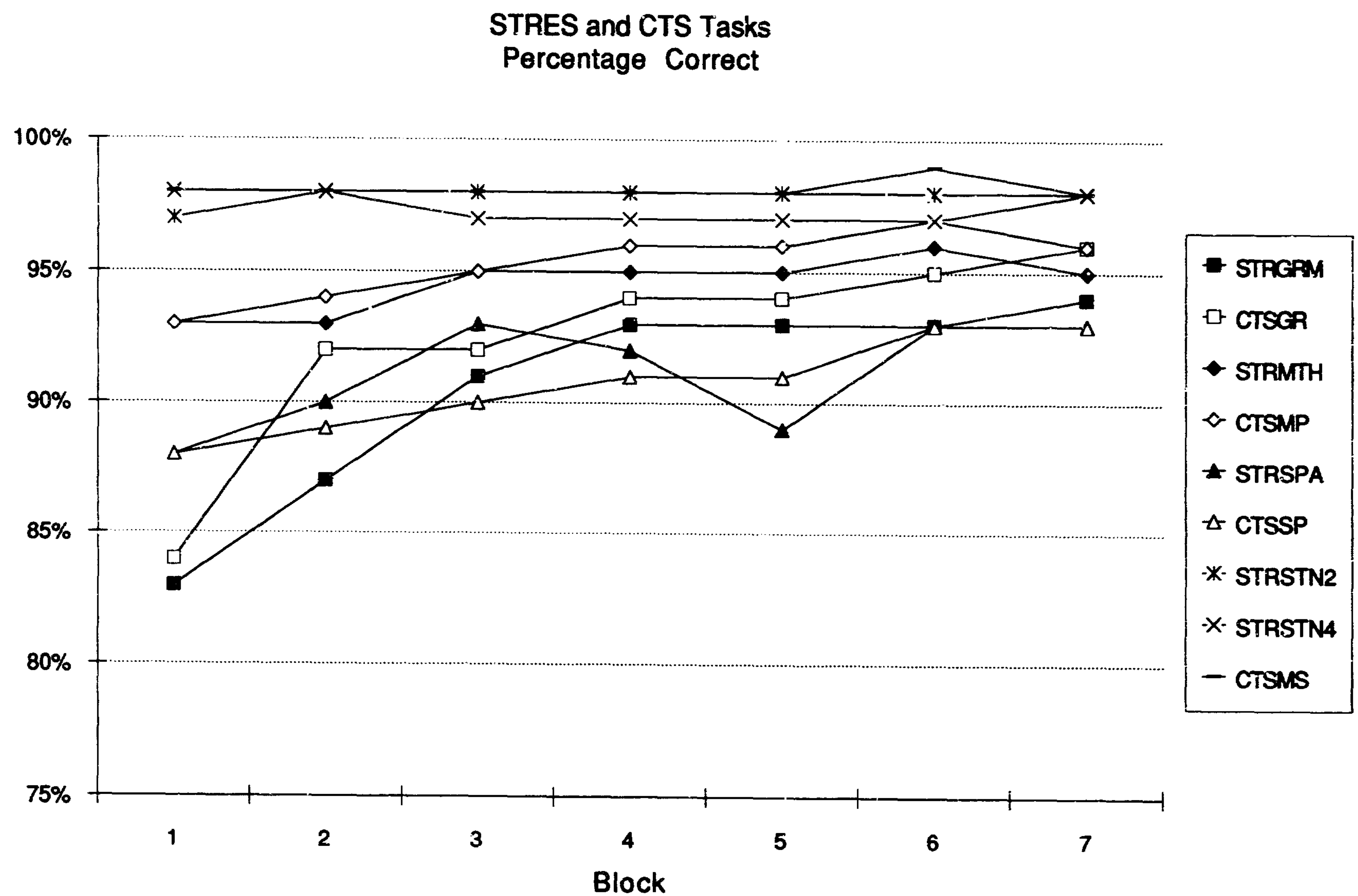


Figure 21. Mean Percentage Correct Comparison of STRES and CTS Discrete Response Tasks.

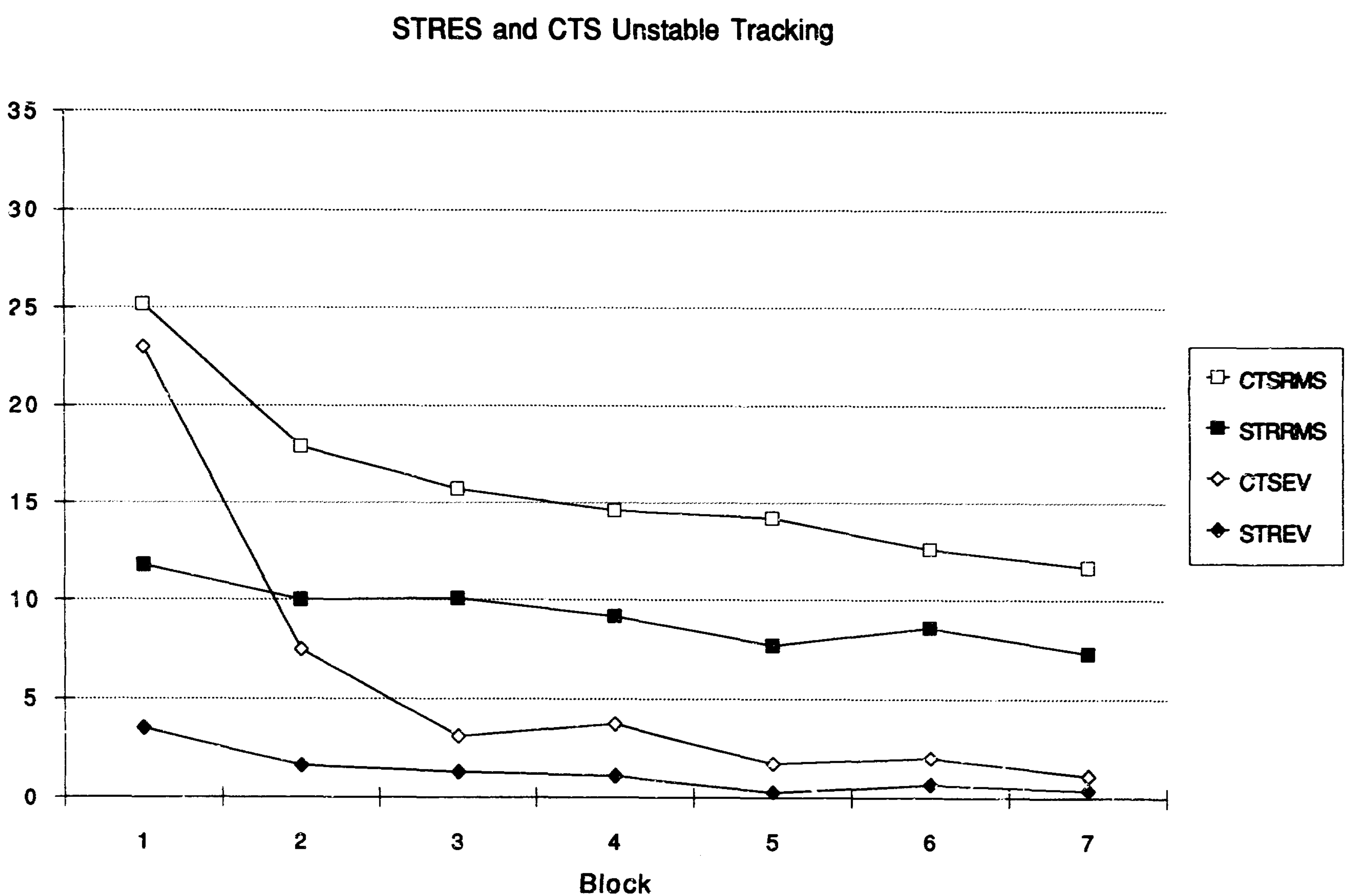


Figure 22. Comparison of STRES and CTS Unstable Tracking Performance Measures.

Response times on the CTS 4-character Memory Search task averaged (60 msec) slower than on the STRES version ($F_{1,78} = 88.80, p = 0.0001$) across all blocks. This statistically significant difference is small and only noticeable when graphed on an expanded scale (Appendix G). However, it is consistent and may represent differences in the hardware characteristics (response keypad vs. keyboard) or software features (display scanning, timing routines) of the two implementations. A statistically significant difference in the Percentage Correct measure ($F_{1,78} = 15.62, p = 0.0002$) was of no practical significance (CTS - 98.2% vs. STRES - 97.5%).

For Spatial Processing, the CTS implementation yielded faster RT's ($F_{1,78} = 249.17, p = 0.0001$). This is likely to be attributable to some combination of (1) differences in the size and spacing of the histogram bars, (2) filled (CTS) vs. outlined unfilled (STRES) bars, and (3) implementation differences in terms of what constitutes a comparison stimulus that is "DIFFERENT" from the standard stimulus. In STRES, a difference of a single unit for one bar is sometimes the only difference between the standard and comparison stimuli whereas in the CTS, the difference is more noticeable on all trials. Note that there is no difference in the percentage correct measure for the two implementations.

Examining only the University of Oklahoma data, substantial differences existed in the performance measures for the two tracking tasks with the STRES version yielding better performance in terms of Edge Violations ($F_{1,62} = 42.44, p =.0.0001$) and RMS Error ($F_{1,62} = 123.57, p = 0.0001$). These differences exhibit both statistical and practical significance and point to the difficulty of implementing a continuous analog task on two different systems. Here again, the differences are likely to be a result of hardware differences (rotary controller vs. joystick, analog-to-digital converter characteristics) and software differences (display resolution and appearance, control loop programming and gain). It is not clear how this problem could be remedied or how a system could be calibrated to yield performance consistent with performance on other systems.

An alternative method of presenting the CTS-STRES comparison is to plot data from all training and baseline blocks on a day-by-day basis. Graphs plotted in this fashion are also provided in Appendix G. Although the correspondence between the CTS and STRES implementations is less clear with this approach, these presentations lead to conclusions similar to those discussed in the preceding paragraphs. Statistical analyses corresponding to this presentation method involved comparing CTS performance with the average of the two STRES blocks on a day-to-day basis.

With this approach, the only task which did not yield significant battery differences was Mathematical Processing. Performance on CTS GR averaged 295 msec slower ($F_{1,78} = 29.90, p = 0.0001$) and slightly less accurate ($F_{1,78} = 4.14, p = 0.045$). CTS MS was 60 msec slower with 1% higher accuracy. CTS SP averaged 200 msec faster with 1% lower accuracy.

As found in the block-by-block analysis, the two measures for tracking indicated significant implementation differences with the STRES version remaining the easier task. On the average, the CTS version yielded more than 8 Edge Violations per trial compared with less than one EV for the STRES version.

In summary, Mathematical Processing performance was quite similar for both implementations, Memory Search was slightly (60 msec) yet significantly faster in the STRES implementation, and Spatial Processing was significantly slower in the STRES implementation. The Percentage Correct measure did not differ appreciably between implementations for any of the discrete tasks. Unstable Tracking performance was substantially better in the STRES version.

A final session-by-session comparison of the CTS data, data from the first STRES block and data from the second STRES block each day shows a consistent improvement from the first to the second STRES block with the CTS better, worse, or the same as discussed previously.

It must be emphasized that due to the relatively large number of subjects tested and the corresponding high statistical power, the statistical significance of the differences between the batteries far outweighs their practical significance. The differences in RT ranged from 60 to 300 msec with typically less than a 1% difference in accuracy. These differences represent a range that is substantially less than the improvements that take place from trial to trial over the course of training. In essence, the agreement between the batteries is substantial. This is further illustrated by the correlation coefficients presented in Tables 23 and 24.

A comparison of the current CTS and STRES database with CTS data from a previous large-scale normative CTS study (Schlegel and Gilliland, 1990) shows substantial agreement for some tasks and disagreement for others (Appendix H). Tasks showing close agreement include Mathematical Processing and Memory Search.

There is some disagreement for the Grammatical Reasoning and Spatial Processing tasks probably pointing to differences in the test samples rather than structural differences in the tasks. This is probably not the case for the Unstable Tracking task where substantial performance differences exist due to task differences (horizontal vs. vertical tracking), software changes in the timing loop subroutine and controller characteristics (potentiometer model and wear).

Table 23. Correlation between CTS and STRES Measures by Block.

Task	Var.	Block 1	2	3	4	5	6	7
GR	RT	0.71	0.80	0.81	0.79	0.76	0.82	0.82
	PC	0.48	0.13	0.39	0.34	0.26	0.09	0.25
MP	RT	0.71	0.79	0.87	0.83	0.76	0.80	0.77
	PC	0.34	0.44	0.42	0.33	0.40	0.39	0.42
MS4	RT	0.74	0.71	0.70	0.62	0.71	0.63	0.61
	PC	-0.06	0.08	0.40	0.25	0.18	0.24	0.21
SP	RT	0.64	0.78	0.76	0.77	0.79	0.72	0.67
	PC	0.52	0.51	0.09	0.21	0.14	0.14	0.22
UT (OKLA)	EV	0.26	0.15	0.28	0.08	-0.06	0.29	0.03
	RMS	0.53	0.59	0.68	0.55	0.32	0.57	0.41

Table 24. Correlation between CTS and STRES Measures by Day.

Task	Var.	Day 1	2	3	4	5	6	7
GR	RT	0.78	0.89	0.90	0.87	0.84	0.90	0.89
	PC	0.55	0.52	0.66	0.48	0.23	0.62	0.44
MP	RT	0.79	0.86	0.92	0.91	0.89	0.89	0.93
	PC	0.34	0.47	0.67	0.60	0.50	0.35	0.58
MS4	RT	0.74	0.78	0.75	0.68	0.78	0.74	0.67
	PC	0.09	0.31	0.36	0.21	0.31	0.37	0.34
SP	RT	0.67	0.82	0.87	0.88	0.85	0.86	0.87
	PC	0.60	0.43	0.22	0.31	0.44	0.29	0.18
UT (OKLA)	EV	0.37	0.23	0.38	0.76	0.41	0.31	0.32
	RMS	0.57	0.68	0.74	0.68	0.62	0.78	0.62

5.5 Effect of Group vs. Individual Testing Procedures

Group versus individual testing procedure effects were examined by comparing the performance of the Armstrong Laboratory subjects, who were tested individually, to that of the University of Oklahoma subjects, who were tested in groups. Because the University of Oklahoma subjects were tested in two phases (cycles) these subjects were divided into a Phase 1 group and a Phase 2 group for the purposes of this analysis. Figure 23 presents representative data plotted by testing group -- in this case, the mean response time variable for the STRES Mathematical Processing task. As with most other dependent variables, there is extremely good correspondence among the groups with regard to the data plots across testing sessions. Similar graphs for each dependent measure for the tasks of the STRES battery, CTS, and WRAIR PAB are located in Appendix I.

The analyses conducted to examine the group versus individual testing procedures effect consisted of a two-way ANOVA with three levels of testing group (Armstrong Lab, OU-Phase 1, and OU-Phase 2) and Trials (either Baseline or Training) conducted for each dependent variable. This analysis also allowed a comparison of the two testing cycles at the University of Oklahoma, i.e., a comparison of two groups of subjects tested under essentially identical conditions.

Because this overall analysis included numerous tests, it would have been appropriate to adjust the experiment-wise Type I error rate through some procedure such as dividing alpha. Of course, in this case, where one might hope that no differences would emerge among the subject groups, Type I error rate control procedures make the detection of significant differences more difficult -- and thus, work in favor of finding fewer significant differences. A statistically less conservative approach was taken in the analyses of this project. Alpha was set at $p \leq 0.01$ to provide a conservative Type I error rate, but not divided further. Thus, the probability of identifying significant differences was sharply increased. This less conservative approach was considered acceptable given that even small differences were viewed as important.

The analyses of Baseline data yielded only a few significant differences. First, significant differences with respect to Testing Procedure were found for both Edge Violations ($p < 0.0001$) and RMS Error ($p < 0.0001$) on the CTS Unstable Tracking task (Figure 24). For both measures, the differences revealed that the Armstrong Laboratory group differed from the two University of Oklahoma groups. This was caused by the controller problem discussed in Section 5.1.3 and resulted in dramatically different scores for the Armstrong Laboratory subjects. These differences were simply a result of this equipment problem.

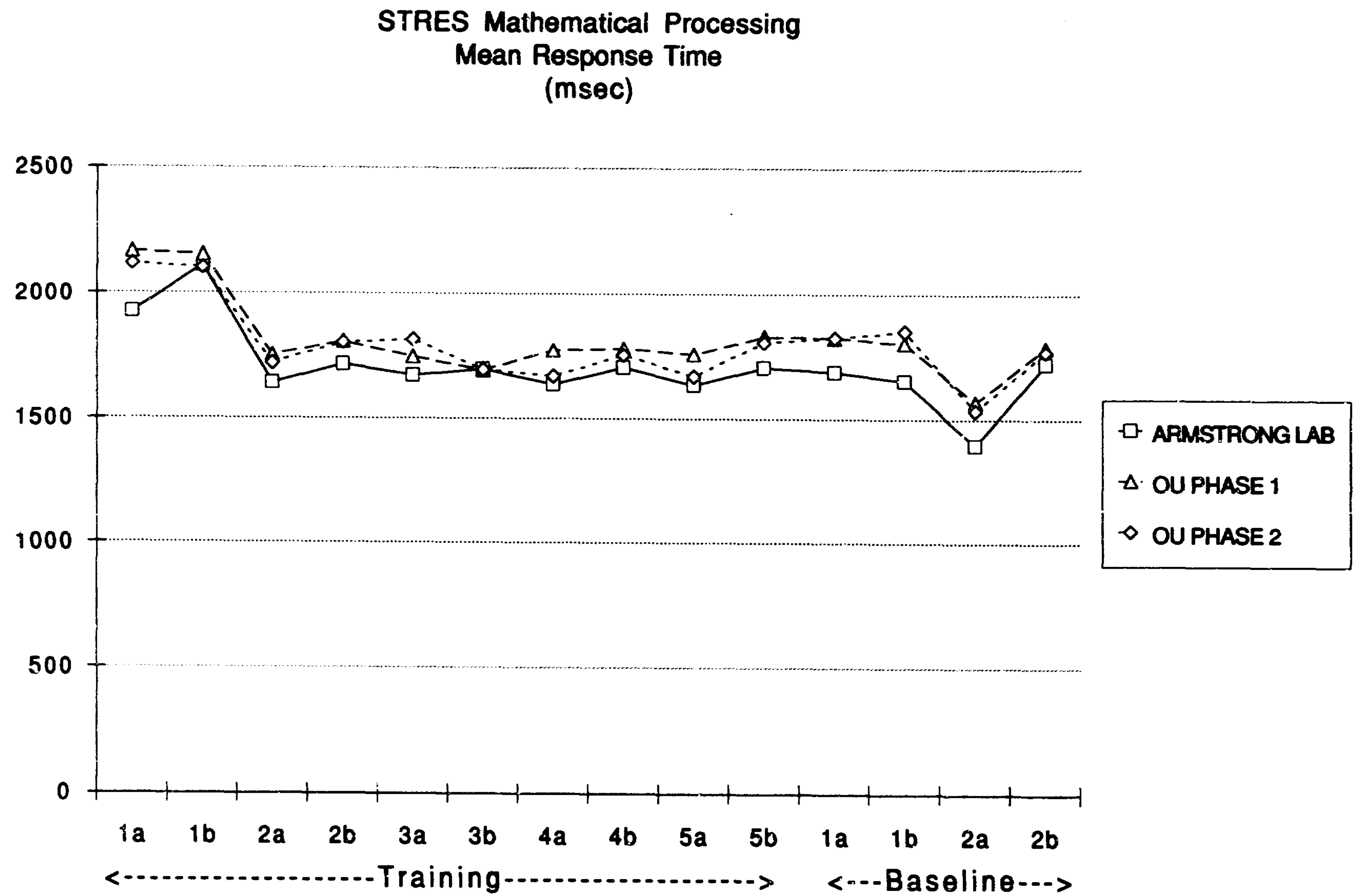


Figure 23. Group vs. Individual Test Administration Procedures for STRES Math Processing.

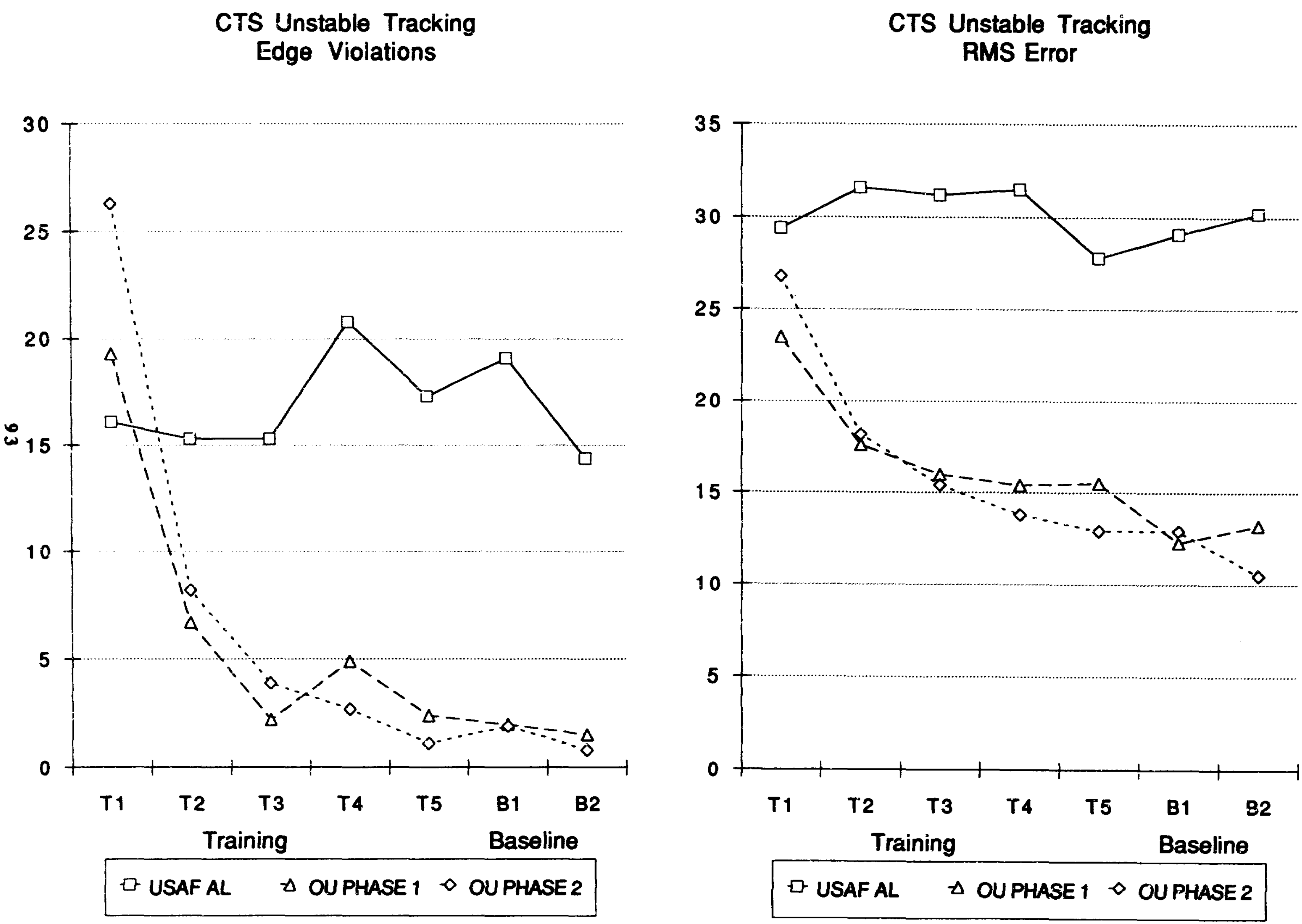


Figure 24. Group vs. Individual Administration Procedures for CTS Unstable Tracking.

Only two additional differences were found and these were marginally significant. The analyses yielded a significant difference between testing groups for the percentage correct measure of the STRES Grammatical Reasoning task ($p = 0.013$) and for the percentage correct measure of the STRES Spatial Processing task ($p = 0.015$). Neither of these analyses actually reached the critical p value and subsequent Tukey multiple comparison tests failed to detect any differences at the 0.01 level. Thus, it was concluded that these differences were inconsequential.

A second similar set of ANOVA analyses were conducted for the Training data. As expected, these analyses also yielded significant differences for Edge Violations ($p = 0.007$) and RMS Error ($p < 0.0001$) for the CTS Unstable Tracking task due to the controller problem. Significant differences were also found for two STRES Reaction Time task forms. Analyses of percentage correct for the first Basic series ($p = 0.0003$) yielded a significant difference and the analyses of percentage correct for the final Basic series ($p = 0.036$) yielded a marginally significant difference. Subsequent Tukey multiple comparison tests yielded no differences at the 0.01 level, but did yield significant differences at alpha = 0.05 between the Armstrong Laboratory subjects and the Phase 1 University of Oklahoma group. The Phase 2 subject performance was closer to that of the Armstrong Laboratory subjects. This difference was traced to a procedural difference in which Phase 1 subjects at the University of Oklahoma had not been exposed to the Basic Reaction Time task during the Orientation session. While their data were affected on the first training day, this difference did not exist on subsequent days.

The only other differences that were identified by the ANOVA analyses were marginally significant differences for the percentage correct measure of the STRES Spatial Processing task ($p = 0.02$) and the RMS Error measure for the STRES Unstable Tracking task ($p = 0.05$). Subsequent Tukey multiple comparison tests failed to yield any significant differences at the 0.01 level. Thus, these findings were viewed as inconsequential. These analyses suggest that the type of training that subjects received had no influence on performance. That the analyses did detect two training/testing related problems (i.e., controller failure and missed orientation) indicates that this analysis was sensitive to variables that would affect performance. However, there was no evidence for an individual vs. group training effect of the type investigated in this project.

5.6 Effects of Task Order and Battery Sequence

While prudence would always suggest counterbalancing the presentation of tasks or batteries, certain environmental, time, or resource constraints can often partially or completely compromise this safeguard. In such cases where counterbalanced presentation is difficult or impossible, it may

be helpful to understand more accurately the potential of such variables to influence performance data.

As noted in Section 4.6.4, four battery sequences and four task sequences were used by different groups of subjects to examine the effects that the order of presentation of batteries and tasks within batteries might have on task performance. Two general analyses were performed to investigate these questions. The first analysis, performed on the Baseline data, consisted of a set of two-way ANOVA's for Battery Sequence (four levels) and another set for Task Sequence (four levels). As with the testing group analysis, Trial was included as a within-subjects factor. Also, the same approach was used with respect to the establishment of the Type I error level. Alpha was set at $p \leq 0.01$ to provide a conservative Type I error rate, but not divided further. Thus, the probability of detecting significant differences was sharply increased, but this less conservative approach was considered acceptable given that even small differences due to Battery Sequence or Task Order were viewed as important.

Figure 25 provides a representative graph of task performance data plotted by Battery Sequence group. As can be seen, the four sequences corresponded exceedingly well across all testing sessions. Additional graphs for each dependent measure for each task battery plotted by Battery Sequence group are located in Appendix J. The analysis of Battery Sequence across all sessions (i.e., training and baseline) yielded no significant differences at alpha = 0.01. Three analyses (mean response time, $p = 0.04$, and percentage correct, $p = 0.03$, for the STRES Mathematical Processing task, and the mean response time, $p = 0.04$, for the WRAIR PAB Manikin task) yielded differences that only met traditional levels of significance (i.e., alpha = 0.05). Subsequent Tukey multiple comparison tests failed to yield any significant differences even at the $p \leq 0.05$ level.

Figure 26 presents representative data from the CTS Mathematical Processing task plotted by Task Order group. As with the Battery Sequence groups, there is remarkable correspondence among the Task Order groups. Additional graphs for each dependent measure for the tasks in the STRES battery and the CTS plotted by Task Order group are located in Appendix K. The analyses for Task Order yielded absolutely no significant difference for any variable for any task. No analyses were even marginally significant.

Together, these analyses revealed no evidence of Battery Sequence or Task Order effects on the dependent measures of the task batteries. While counterbalancing is always recommended, these data suggest that in cases where it is not possible, Battery Sequence and Task Order effects for the tasks assessed in this project may not pose serious threats.

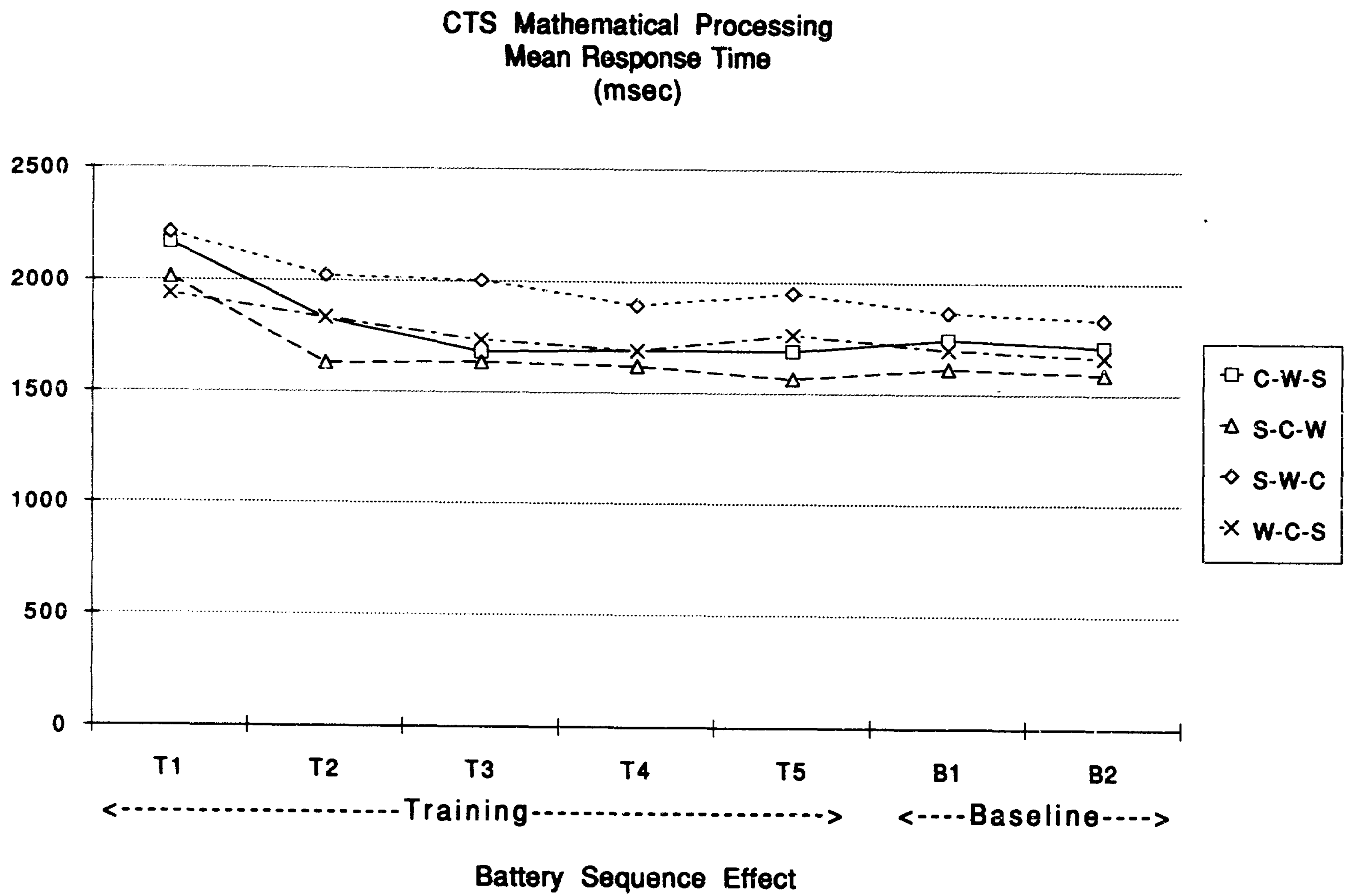


Figure 25. Battery Sequence Effects for CTS Mathematical Processing.

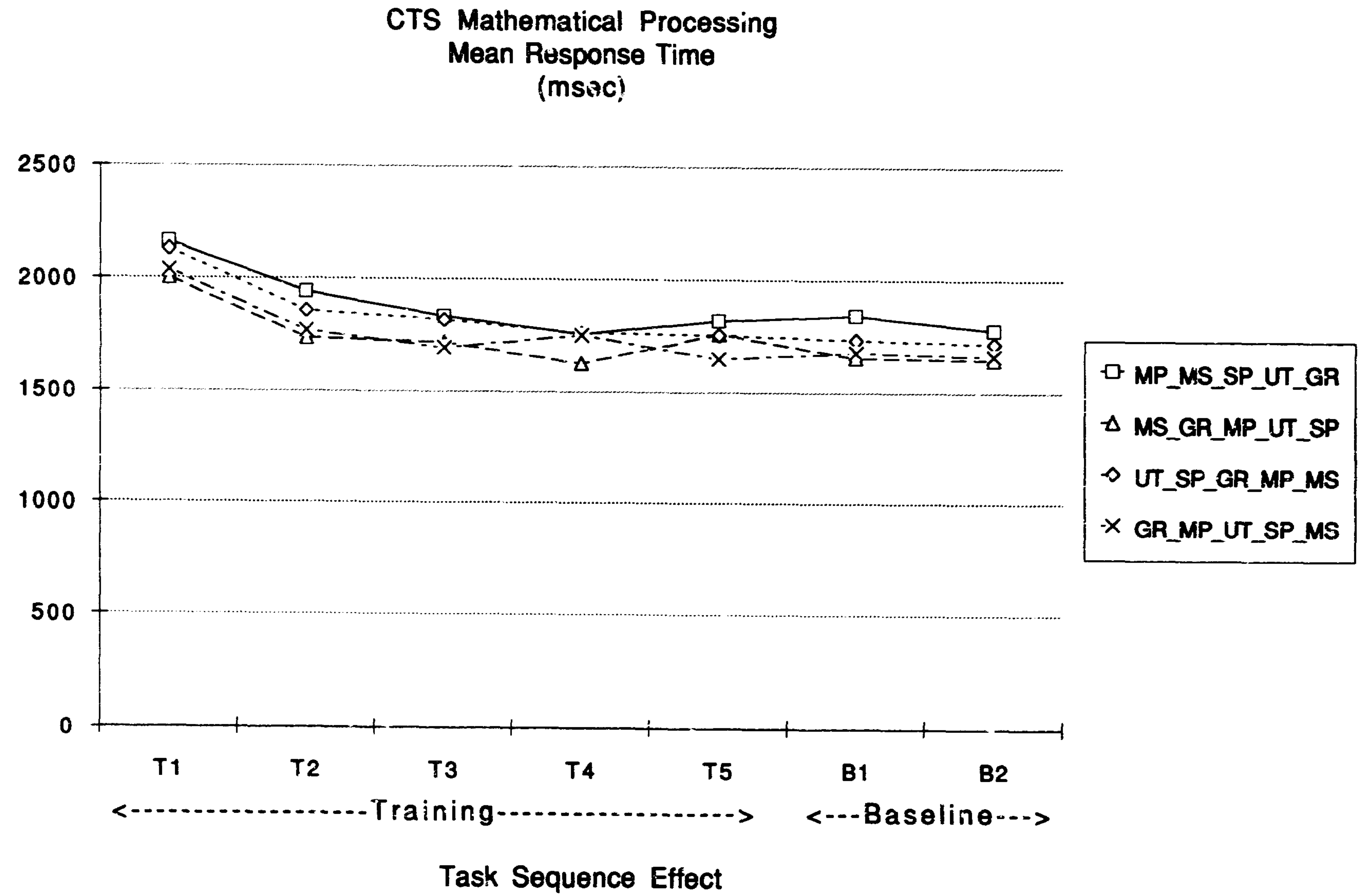


Figure 26. Task Order Effects for CTS Mathematical Processing.

5.7 Effects of Imposing Response Deadlines

Figure 27 presents an example of the effect of imposing deadlines on the responses to discrete tasks, in this case the Grammatical reasoning task from the STRES battery. The term "actual" deadline refers to the situation when an actual deadline was imposed by blanking the CRT display and locking out any response after the deadline. For "pseudo" deadlines, the subjects were presented with instructions indicating the level of deadline imposed, but were not actually restricted in their responses any differently than in the no deadline condition.

The first and most obvious conclusion is that an actual deadline hastened the subject's responses (or did not count the slower responses as acceptable) and resulted in a shorter mean response time for correct responses. The magnitude of the speed increase is related to the severity of the deadline. Accompanying the increase in speed is an increase in the number of totally missed responses, and therefore a decrease in the percentage correct. What may be amazing is the apparent fact that the use of a pseudo deadline has a similar impact in that response time decreases to comparable levels. However, because there is no actual time-out on the responses, there are fewer missed responses, and the percentage correct is essentially the same as for the No Deadline condition.

This is particularly apparent when examining the last series of trials where each subject, regardless of the deadline under which training was conducted produced a similar pattern. Both response time and percentage correct decreased with stricter actual deadlines. Under the pseudo deadline conditions, however, subjects performed faster with no apparent loss in accuracy. This finding is important in terms of developing techniques to maximize subject performance, but adds another complication to determining the reliability of the tasks. A similar pattern of results was obtained with the other four STRES tasks used in the Deadline Study.

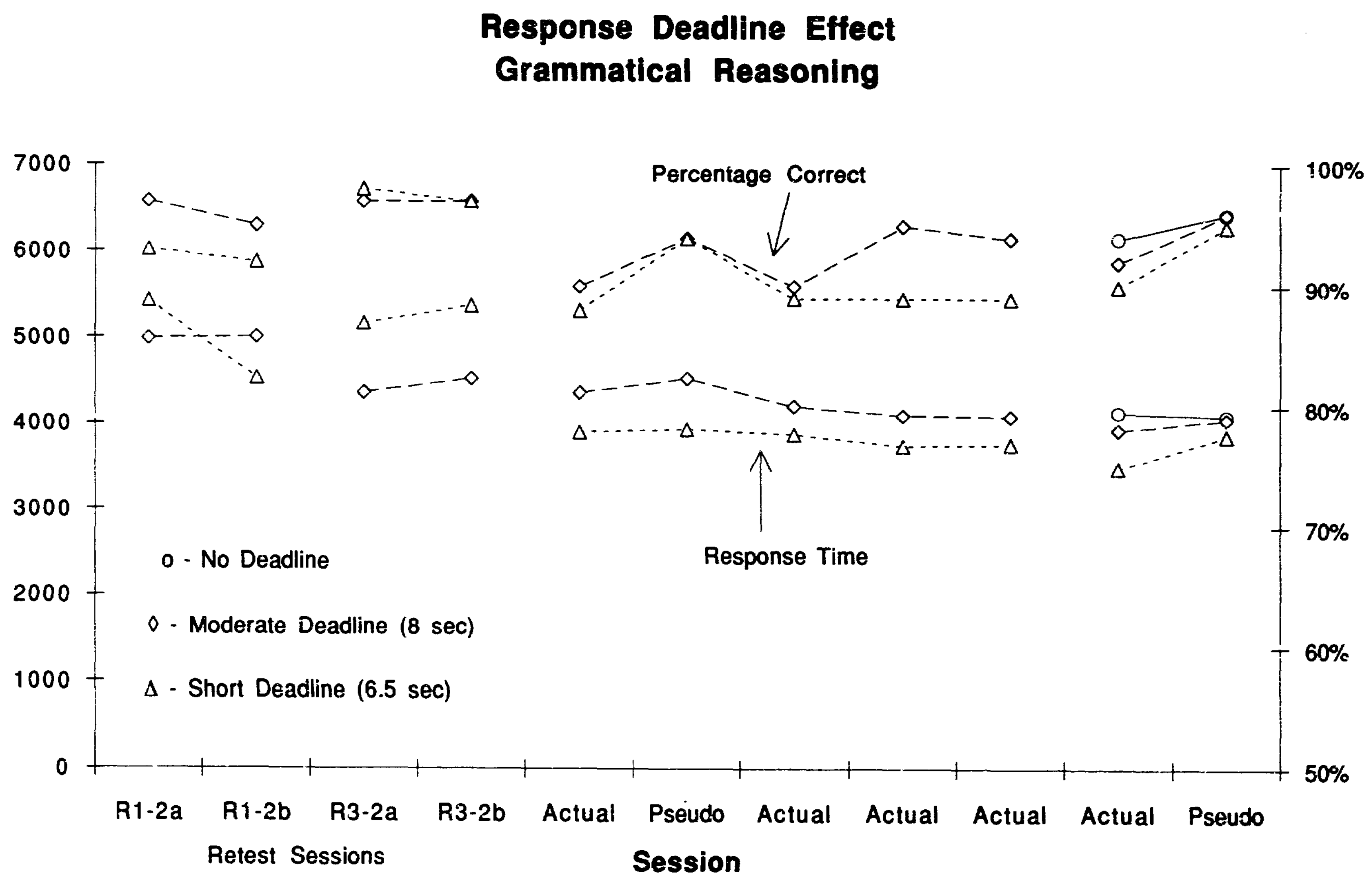


Figure 27. Response Deadline Effect for Grammatical Reasoning Task.

5.8 Effects of Extended Trial Length

The data from the extended trial analyses provide the opportunity to explore the influence of this variable only at a rudimentary level. The fact that (1) protocol limitations did not allow all combinations of the tasks to be offered, and (2) not all trial lengths could be randomly presented, may have compromised the full generalizability of these results. In fact, a visual examination of the data suggest that in some cases the 6-minute trial length data may have been highly affected by such factors. It also appears that the 24-minute data probably provided the best opportunity to explore trial length effects. Nonetheless, the data did provide a preliminary opportunity to examine the influence of this variable.

As was noted in Section 4.0 (Project Design and Method), the extended trial length data were collected on a limited set of STRES battery tasks only. In addition, analysis priority was given to the STRES Unstable Tracking task due to the continuous nature of the task. It was hypothesized that a task that is relatively continuous in nature would not provide brief periods between discrete trials that could be utilized as rest pauses. This constant demand for performance would more likely provide the opportunity to observe any effects of extended trial length on performance. For this reason, the discussion of extended trial length effects will focus on the STRES Unstable Tracking data. However, a more limited examination of results from other tasks will also be presented.

Figures 28 and 29 present the Edge Violation and RMS Error measures, respectively, for the STRES Unstable Tracking task across the various extended trial lengths. The 3-minute trial length data (Baseline Day 2, Trial 1) are presented in the dark column in the foreground followed by each increasing trial length condition divided into 3-minute epochs. It should be noted that 6-minute trial length data (transparent in the figures) appear somewhat unusual and may reflect the fact that this condition was not counterbalanced in its presentation. Due to protocol restrictions, the 6-minute trial length condition always followed one of the longer trial length conditions. Thus, these data may be revealing a fatigue factor that is not evident in other data. Also, because the tracking task is continuous and requires constant motor performance, this fatigue effect may be more likely to emerge in these data than perhaps in any other task.

Collapsing the data across all 3-minute epochs within each trial length condition provided the opportunity to examine whether there was a difference in average performance across the four trial lengths. This one-way ANOVA resulted in a significant trial length effect for RMS Error ($F_{3,90}$ = 8.99, $p < 0.0001$), but only a marginally significant effect for Edge Violations ($F_{3,90}$ = 3.15,

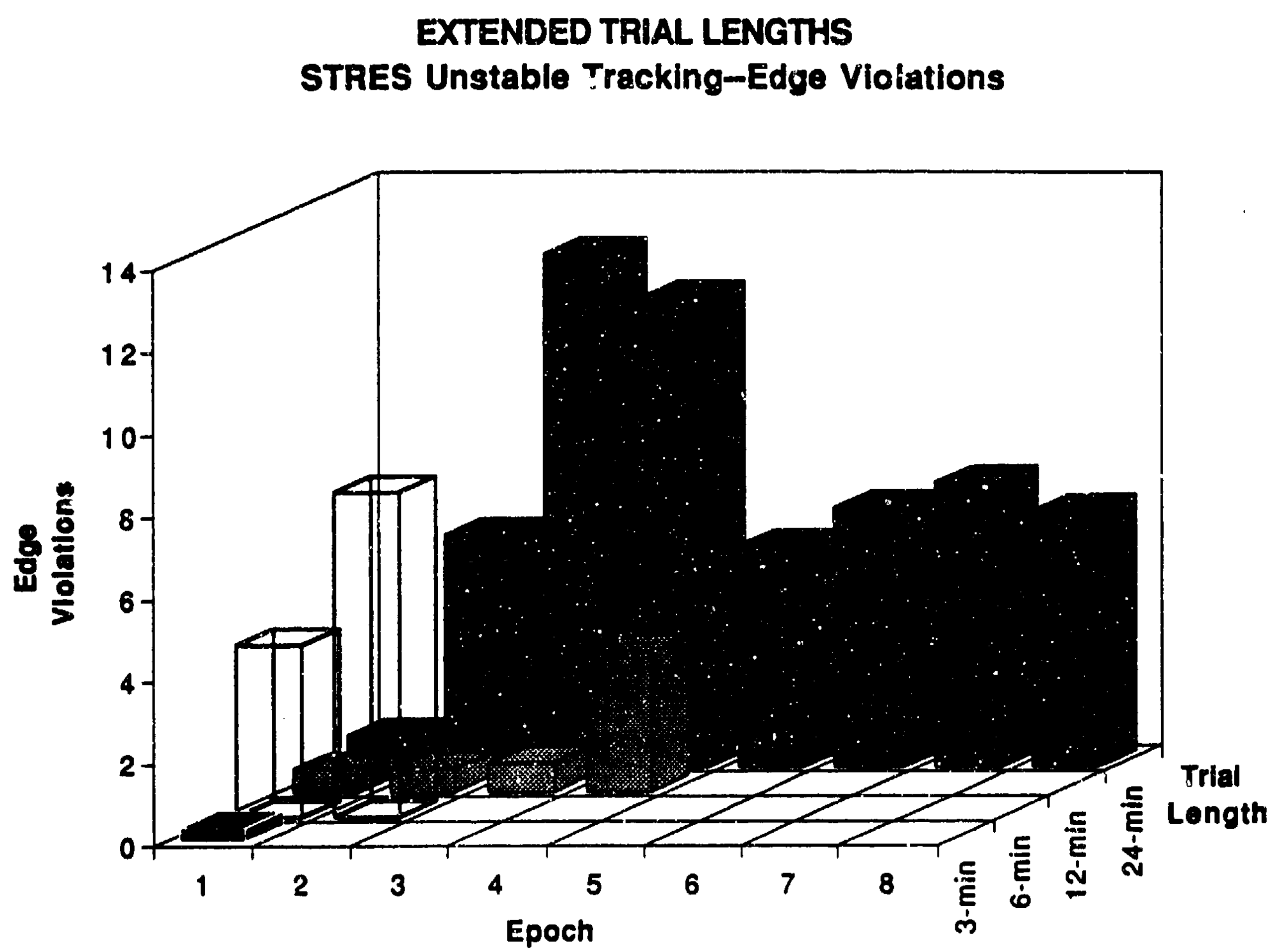


Figure 28. Extended Trial Length Data - Unstable Tracking Edge Violations.

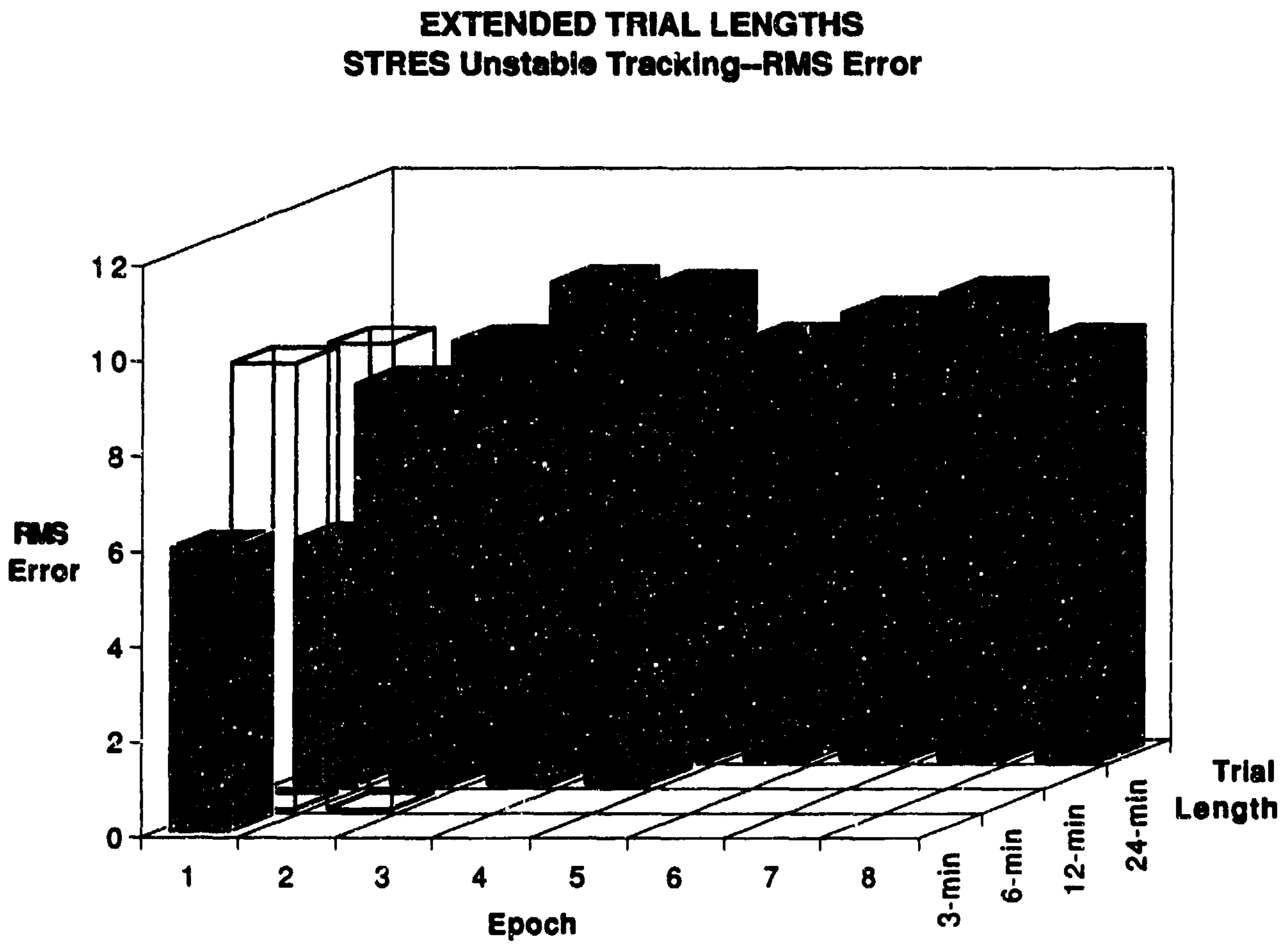


Figure 29. Extended Trial Length Data - Unstable Tracking RMS Error.

$p = 0.07$).[3] Subsequent multiple comparisons (paired t-tests corrected for experiment-wise type I error rate) revealed that the Baseline 3-minute trial differed from the 6-minute trial length condition ($t_{30} = 3.85$, $p = 0.0006$) and the 24-minute trial length condition ($t_{30} = 3.35$, $p = 0.002$). In general, these results based on data collapsed across epochs within the trial lengths suggest that the average Unstable Tracking performance level across the various trial lengths is not the same.

Another one-way ANOVA was performed on the first 3-minute epoch from each trial length condition. This analysis tested whether the inital performance of the subject was different given the various trial length conditions. This analysis yielded a significant trial length main effect for both RMS Error ($F_{3,90} = 9.79$, $p < 0.0001$) and Edge Violations ($F_{3,90} = 7.87$, $p = 0.005$). Subsequent multiple comparisons (paired t-tests protected for experiment-wise error rate) revealed that the only significant differences involved comparisons with the 6-minute trial length data. Due to the somewhat questionable nature of the 6-minute data, it was concluded that, while statistically significant, these were probably not theoretically important findings. Aside from the obviously anomalous data from the 6-minute trial length condition, there appeared to be no significant differences between the first 3-minute epochs in any trial length. Thus, subjects appeared to begin each trial length with fairly comparable Unstable Tracking performance, at least in the first three minutes.

Given that the average Unstable Tracking RMS Error performance of subjects was different across the various trial lengths, another logical question was to explore where these trials differed. The Baseline 3-minute trial was compared to each set of epochs from the other trial length conditions. From the analysis of initial 3-minute epochs, it was clear that RMS Error performance during the Baseline 3-minute trial differed significantly from the initial 3-minute epoch of the 6-minute trial length condition. Not surprisingly, a subsequent analysis confirmed that RMS Error performance from the Baseline 3-minute trial was also significantly better than the second 3-minute trial of the 6-minute trial length condition ($t_{30} = 3.92$, $p = 0.0005$). An analysis comparing the Baseline 3-minute trial to each additional epoch in the 12-minute trial length condition failed to yield any significant differences for RMS Error. Comparisons between the Baseline 3-minute trial and other epochs in the 24-minute trial length condition yielded a number of significant differences for RMS Error performance. In fact, simple paired t-test comparisons of the Baseline trial with all 3-minute epochs other than the first were significant at or near the $p \leq 0.01$ level of significance. However, when alpha was adjusted to protect the family-wise Type I error rate, RMS Error performance during the Baseline 3-minute trial was significantly better than only the third

[3] All values of alpha cited for ANOVA analyses employ Greenhouse-Geisser corrections.

($t_{30} = 3.89$, $p = 0.0005$), fourth ($t_{30} = 3.46$, $p = 0.0016$), sixth ($t_{30} = 3.33$, $p = 0.002$), and seventh ($t_{30} = 3.57$, $p = 0.0012$) epochs in the 24-minute trial length condition.

The above results are evidence that performance did vary across the 24-minute period. Subjects appeared to have periods in which they performed as well as they did during the Baseline 3-minute period. Likewise, they also had periods where their performance was significantly worse than the Baseline 3-minute trial. During the 24-minute trial condition, subjects did not appear to have any periods during which their performance was better than the 3-minute trial condition.

Given some of the limitations in the collection of the extended trial data that were mentioned at the beginning of this section, a very cautious examination of the performance of other STRES battery tasks over the various trial lengths was conducted. These analyses were attempted to reveal general trends of interest only. Extensive and detailed analyses were viewed as inappropriate given the aforementioned limitations. In this regard, comparisons between the Baseline 3-minute trial and data collapsed across epochs within the other trial length conditions revealed that only the percentage correct measures for the STRES Mathematical Processing and the STRES Spatial Processing tasks yielded significant results. For Mathematical Processing, the percentage correct went from approximately 98% for the Baseline three-minute trial to approximately 96% for the 24-minute trial while Spatial Processing declined from 96% to 92%. While percentage correct remained constant across epochs within a given trial length, there was a small, yet statistically significant decline in the average percentage correct across trial length conditions. This suggests that as trial length increases, the average percentage correct decreases.

Of more interest were the results of a general analysis of the 24-minute trial length data. These data were viewed as among the most viable. One-way ANOVA's were performed across epochs for each of the major dependent variables for the STRES battery tasks. The results of these analyses revealed significant differences for the response time measures of the STRES Grammatical Reasoning and Mathematical Processing tasks. Figures 30 and 31 present the data for these tasks, respectively. It can be seen that there is a decrease (improvement) in response time across the epochs. This trend can be seen for other trial lengths as well. This apparent improvement was not expected. Also, it should be noted that the 6-minute trial length condition did not show the rather unusual level of difference that appeared in the tracking data.

Explanations for these results are not entirely clear. It is possible that the continuous nature of the tracking task and its constant demand on psychomotor activity leads to a level of neuromuscular fatigue that is not present in the other tasks. In fact, the intermittent and discrete nature of the other tasks, coupled with possible factors such as increased attention over time to skills necessary for

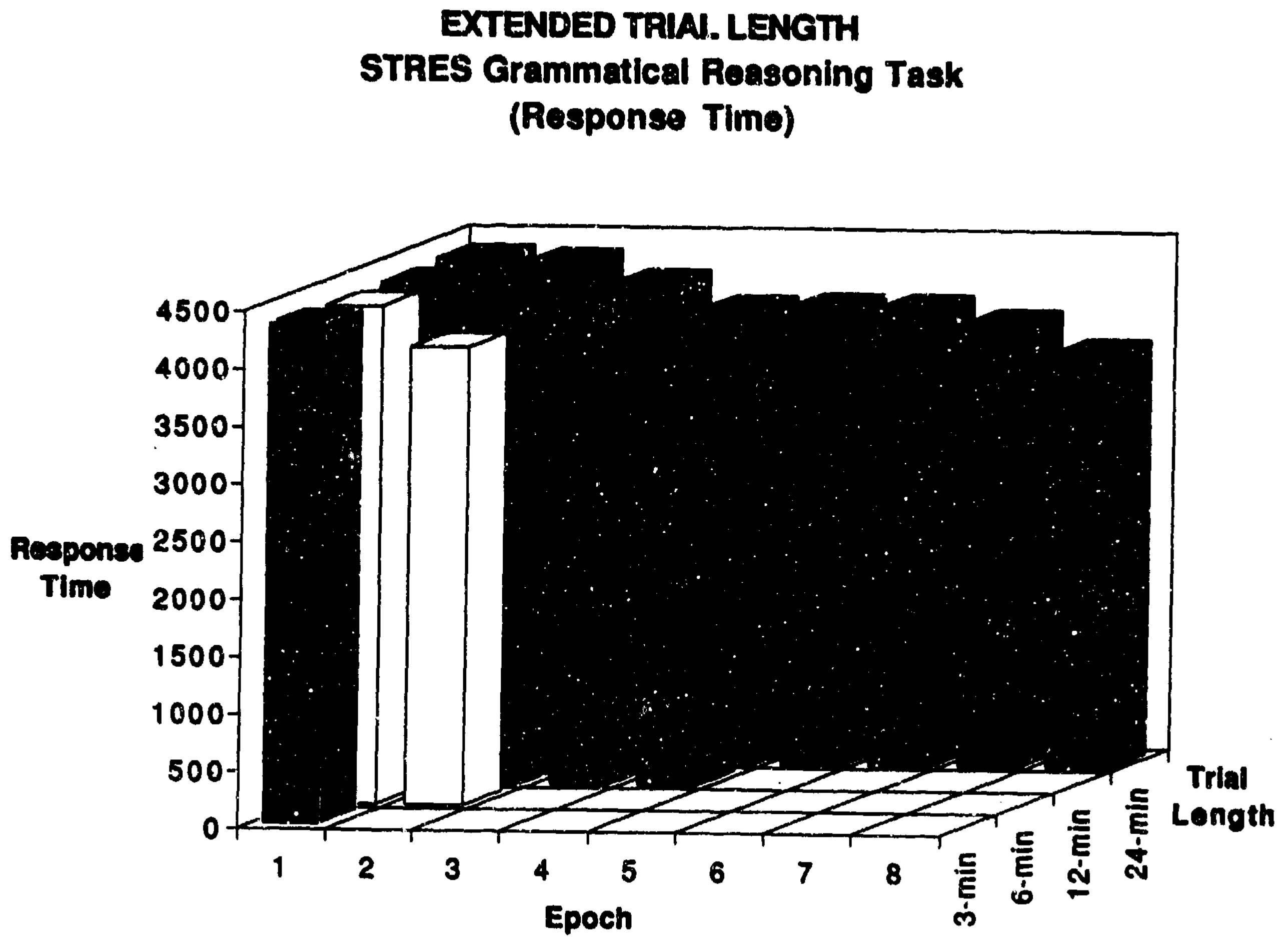


Figure 30. Extended Trial Length Data - Grammatical Reasoning Response Time.

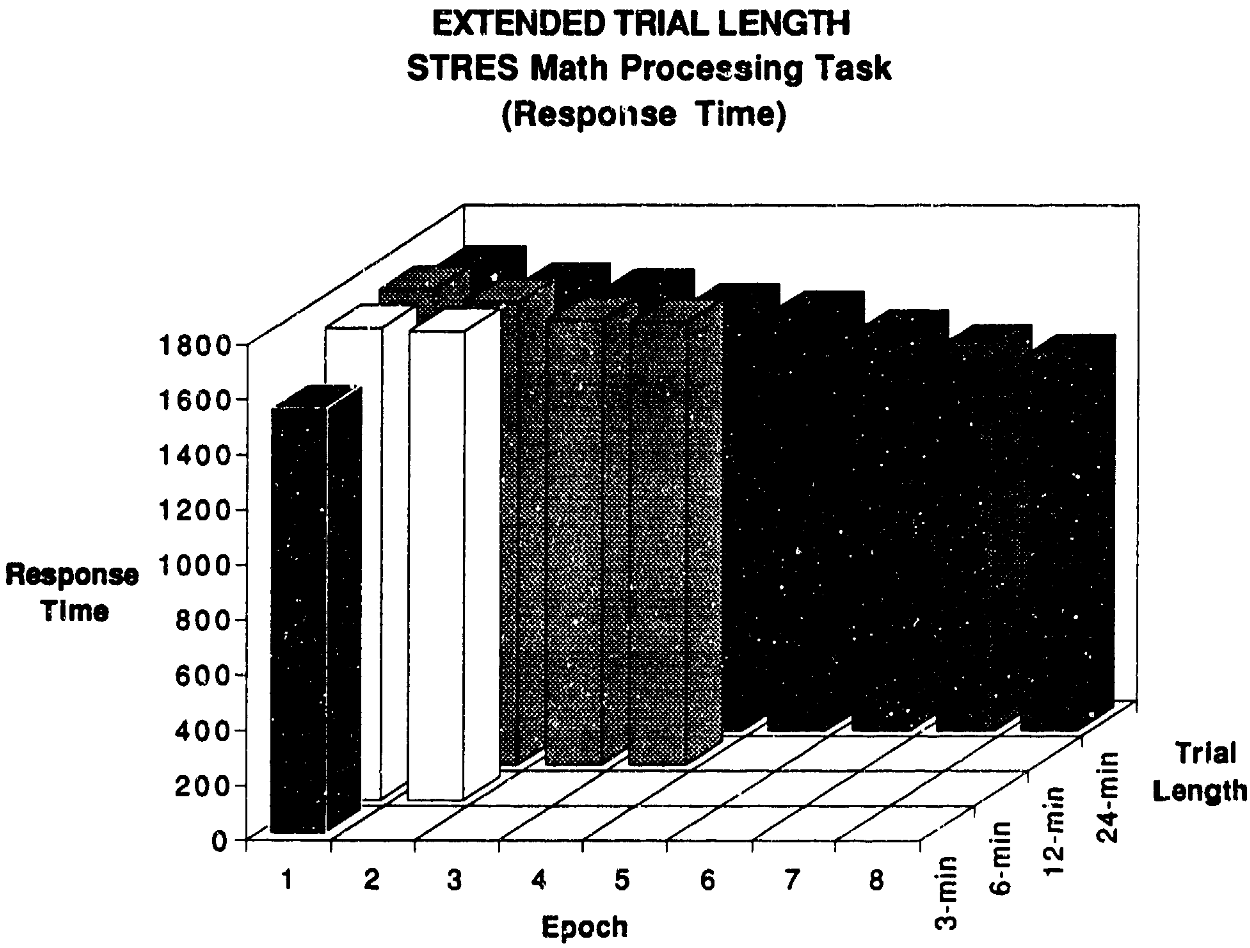


Figure 31. Extended Trial Length Data - Math Processing Response Time.

optimal task performance, may have allowed the subjects to actually become more proficient (respond faster) on these tasks, compared with the tracking task. In other words, continued learning may have taken place with intensified, massed practice. The aberrant nature of the 6-minute trial length tracking data (i.e., the condition most often performed after prolonged tracking trials) gives some support to this hypothesis. It should be noted, however, that the unexpected decrease in response time for these tasks may be derived not only from the opportunity to optimize performance over time, but also from a shift in task strategy. The declines in percentage correct measures noted above may be a subtle indication that the subjects are trading a certain degree of accuracy for gains in speed. This possible regulation of the speed/accuracy tradeoff may be an important index of prolonged performance efficiency.

5.9 Usefulness of Psychometric State Measures

Both the Stanford Sleepiness Scale and the Mood Scale II were administered at various times throughout the testing. These scales were included to obtain information regarding their ease of administration and usefulness as research tools in assessing the impact of testing variables. No extensive attempt was made to perform a detailed analysis of these data at the present time. However, the subjects' responses to these questionnaires were compiled and plotted for visual analysis. It should be remembered that during Training, Baseline, and Retest sessions, these questionnaires were administered prior to and after the subjects completed the STRES battery portion of their daily testing protocol. Half of the subjects in any given testing session performed the STRES battery during the first hour of the two-hour session while the other half of the subjects were performing the CTS-WRAIR PAB tasks. These groups then switched workstations. Thus, half the subjects completed the psychometric questionnaires at the beginning of the testing session and then again at the midpoint in the testing session. The other half of the subjects completed the questionnaires first at the midpoint of the testing session and then after the testing session ended.

Figure 32 presents the responses of subjects on the Stanford Sleepiness scale. The subjects performing the STRES Battery first in the testing session are plotted in the foreground. The other half of the subjects, who performed the STRES battery during the second half of the testing session, are plotted in the background. From this figure, it appears that subjects were generally low in sleepiness throughout the testing period. However, they appeared to increase in sleepiness slightly during the first one hour of testing. The subjects who performed the STRES battery second show a similar level of sleepiness at the midpoint in the testing session as compared to the group performing the STRES battery first. They then reported an additional increase in sleepiness during the final one hour of testing. By examining both the between- and within-group data, it appears that sleepiness gradually increased across the two-hour testing session. It should be

remembered that these trends were not confirmed with statistical tests, yet the trends are clearly confirmed by anecdotal reports. Subjects often reported feeling more tired and fatigued at the end of the sessions.

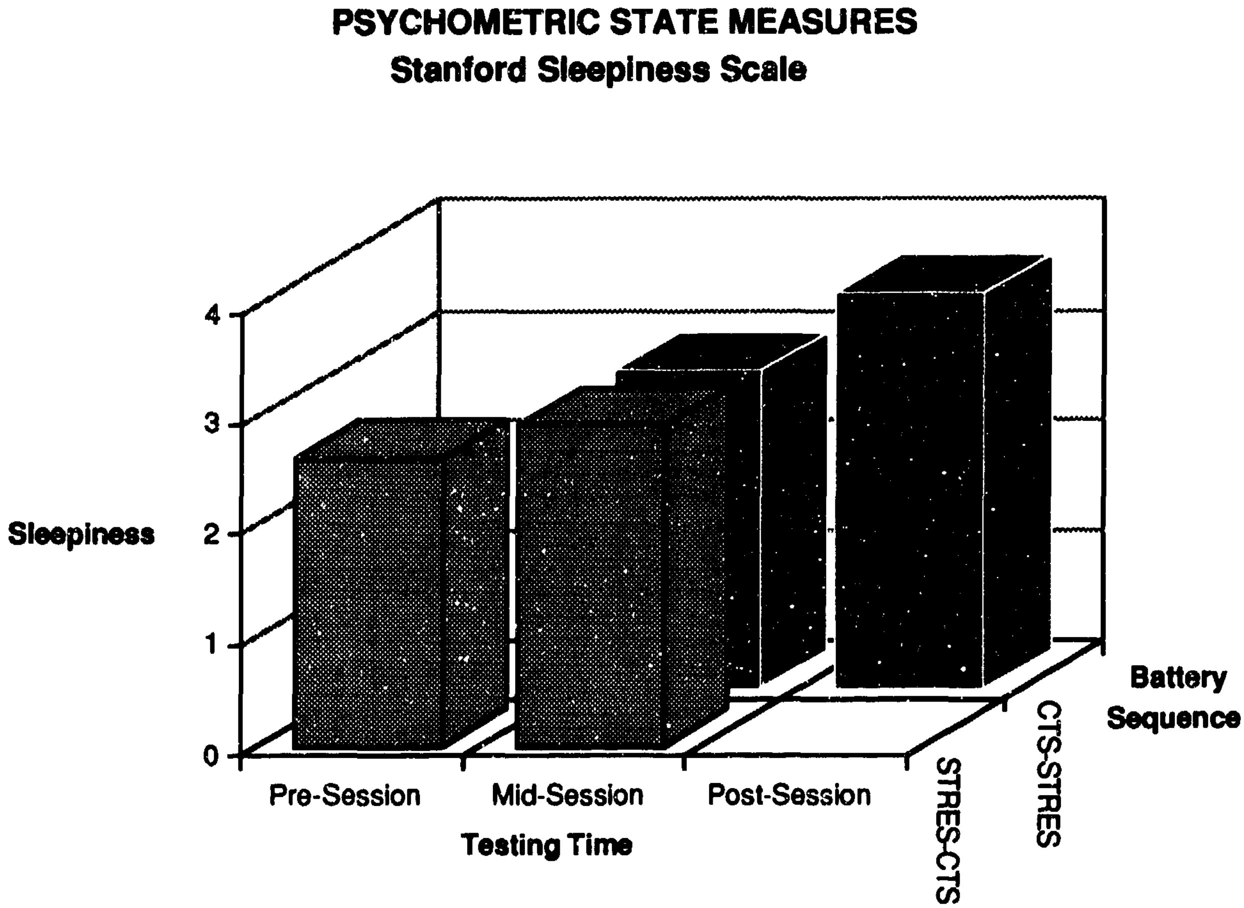


Figure 32. Stanford Sleepiness Scale.

The trends seen in the Stanford Sleepiness Scale were also confirmed, to some degree, by similar trends in relevant scales of the Mood Scale II. Figures 33 through 39 present the various subscales of the Mood Scale II, as well as the mean response time to complete the scale. It appears that the Activity and Happiness subscales demonstrated trends somewhat like the Stanford Sleepiness scale. That is, these two scales revealed modest declines in activity level and happiness through the test session, especially from the first administration to the second administration within the testing session. However, these data did not show the continual decline found in the sleepiness scale.

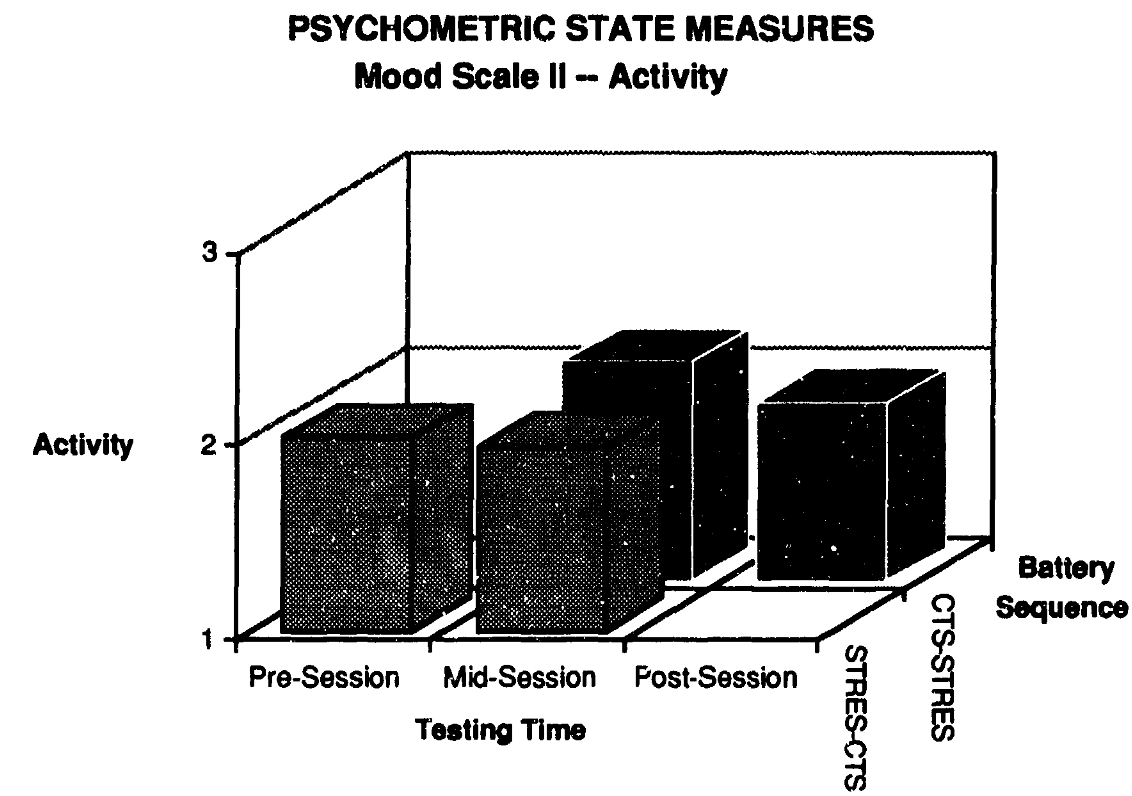


Figure 33. Mood Scale II - Activity Scale.

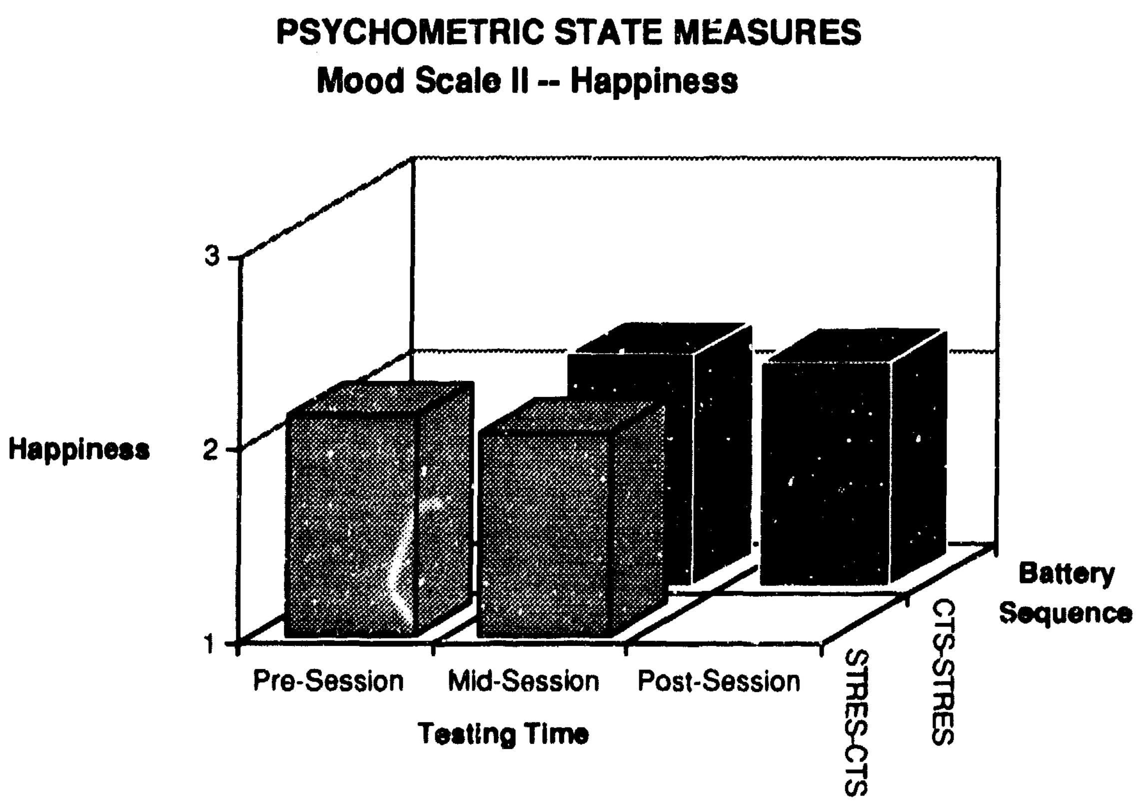


Figure 34. Mood Scale II - Happiness Scale.

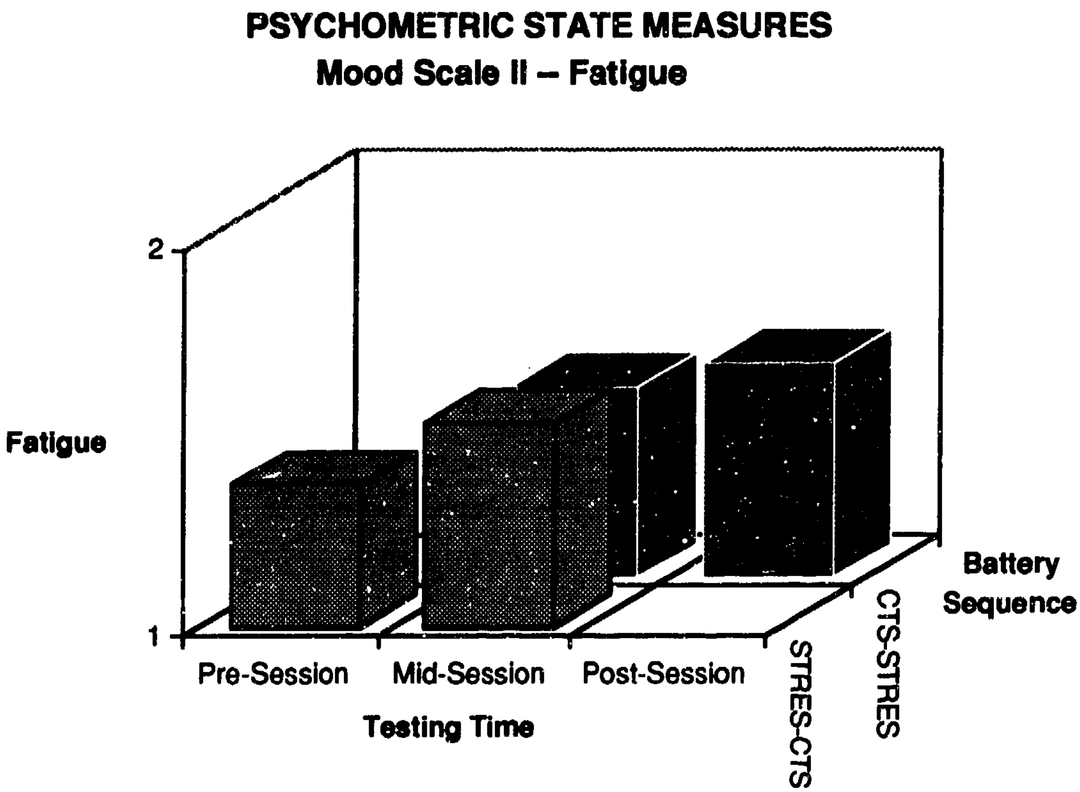


Figure 35. Mood Scale II - Fatigue Scale.

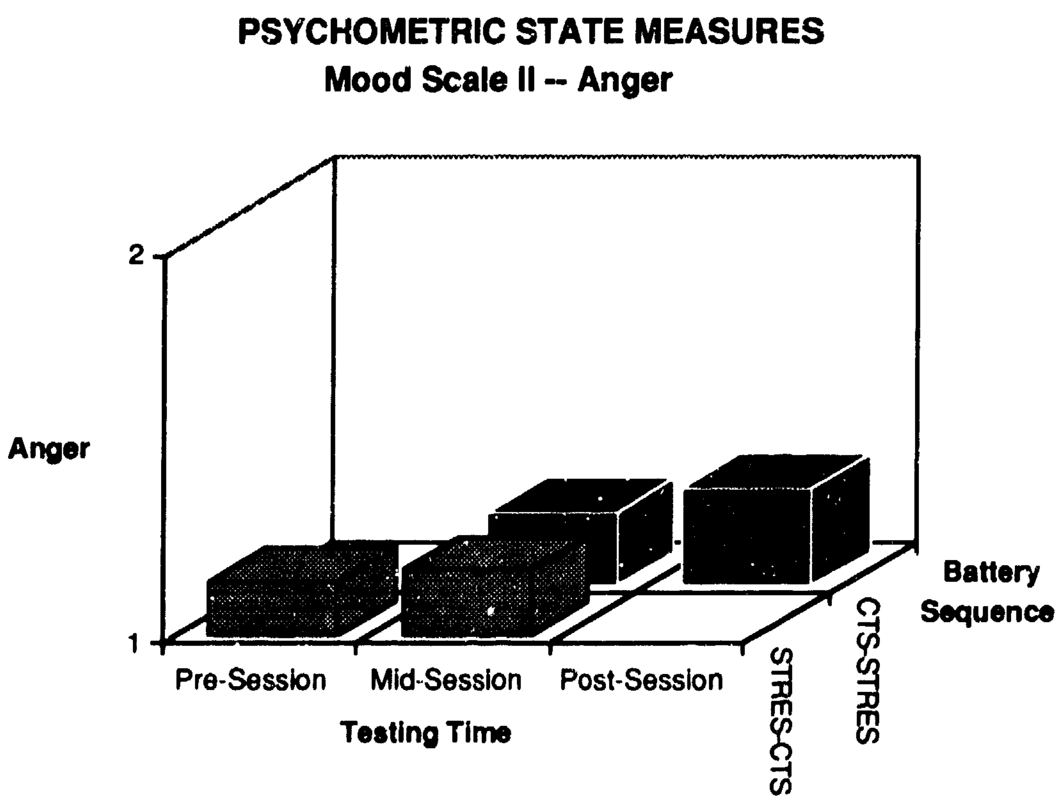


Figure 36. Mood Scale II - Anger Scale.

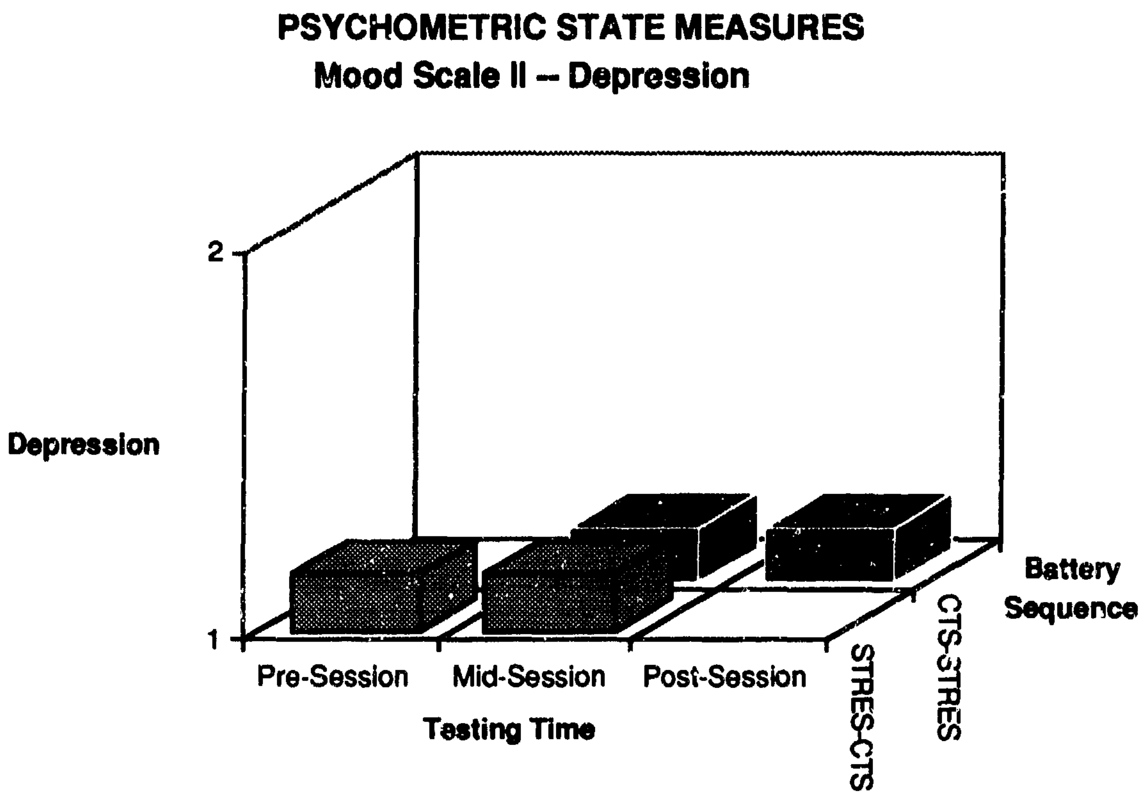


Figure 37. Mood Scale II - Depression Scale.

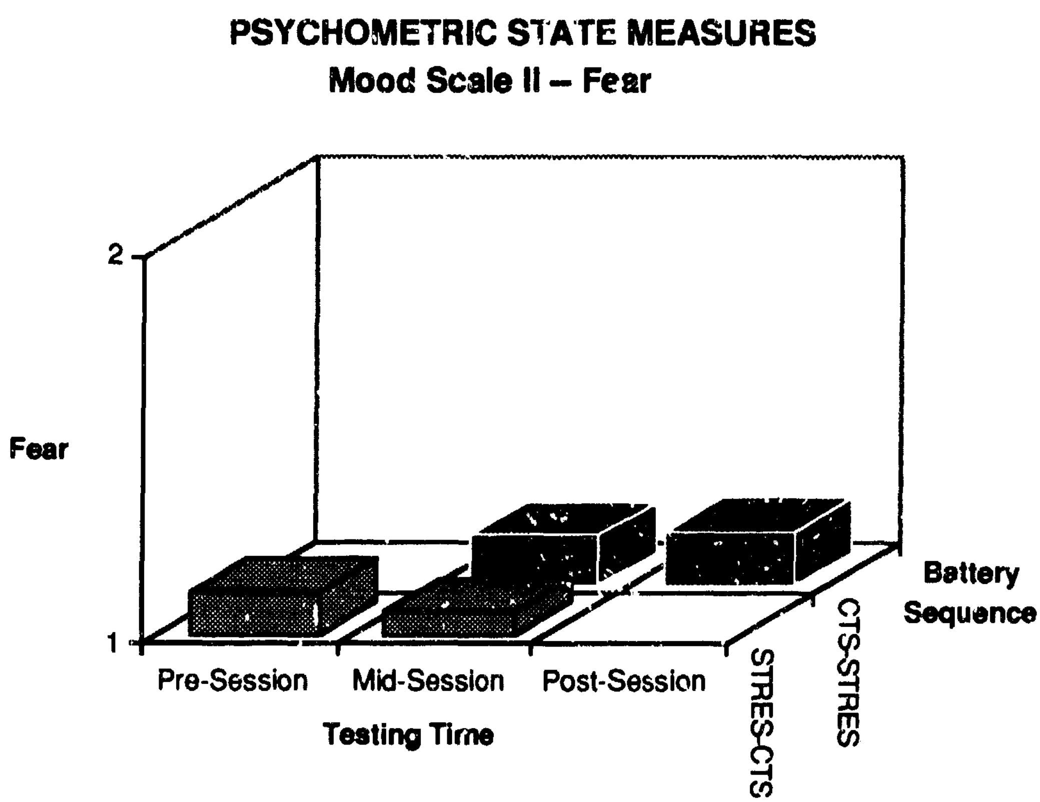


Figure 38. Mood Scale II - Fear Scale.

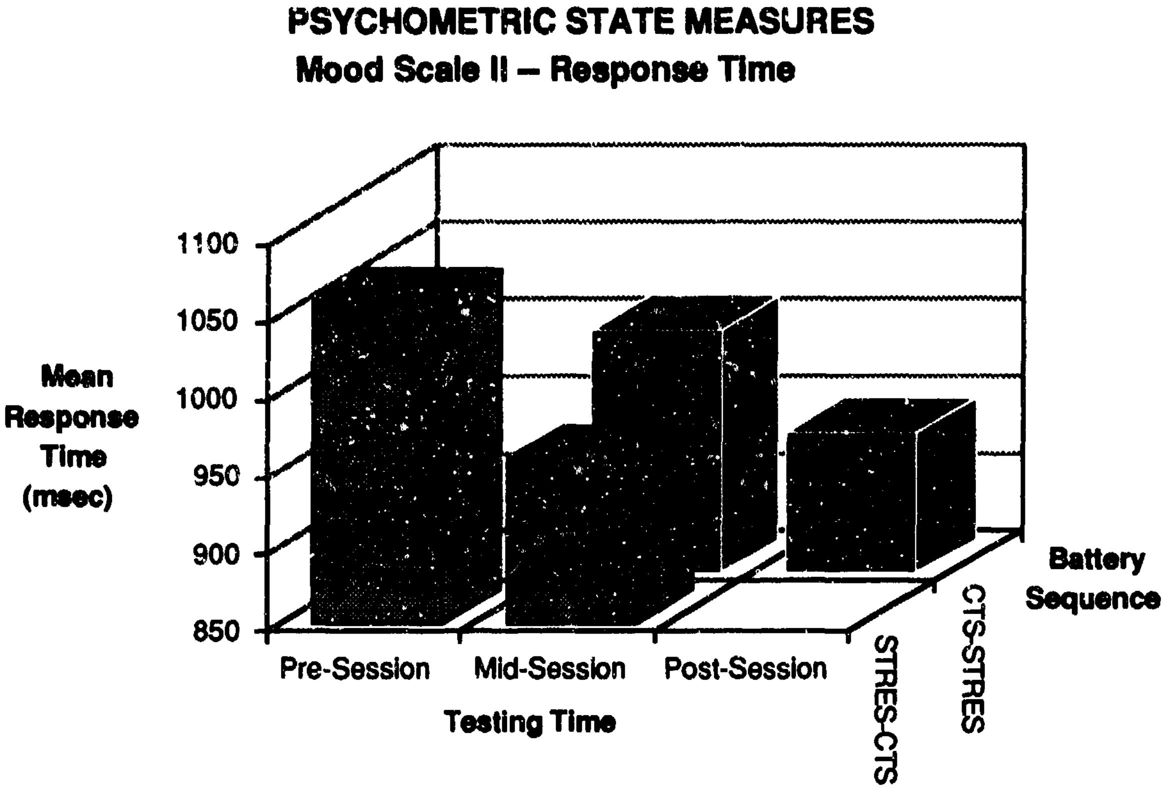


Figure 39. Mood Scale II - Response Latency.

Scales such as Fatigue and Anger (frustration) would be predicted to show trends in the opposite direction. In fact, a visual inspection of these scales reveals that they did indicate apparent increasing levels of self-reported fatigue and anger. Both trends appeared to be fairly continuous across the testing session.

The two more clinically-related scales that would be predicted to have very little relationship to psychological state during testing revealed little to no change across the testing session. Finally, it appears that the subjects completed the questionnaires more slowly during the first administration of the scale in comparison to the second administration during each testing session.

These data appear encouraging with respect to the potential usefulness of psychometric measures in assessing psychological state during task performance. However, several points should be kept in mind. First, this was a very preliminary analysis. Trends in the subjects' responses were clearly evident and the trends were supported by similarities in dimensions that were rationally related. However, these trends were not tested statistically. Second, the range of scores in most cases was very low, many times less than one scale point (out of seven for Stanford Sleepiness and out of three for Mood Scale II). The trends and consistencies that were seen were all the more impressive for this reason, yet adequate measurement ability will probably require

greater use of the full range of the scale score responses. Finally, if a factor was assessed by these data it was probably the cumulative effect of testing over the course of two hours. This effect may have been quite small, which still raises some encouragement for the use of such scales. More powerful effects, such as drugs or environmental variables may be good candidates for further study of the usefulness of these psychometric scales.

6.0 SUMMARY

The following comments summarize the results of this project. While there were numerous detailed analyses conducted during this study, and while there are still many more that can be performed, there are a number of general statements that characterize the major findings of this project.

- A substantial database has been established for selected tasks from the STRES battery, CTS and WRAIR PAB based on a well-defined population and a sizable number of performance trials. Percentile breakpoints at 20% increments were included to allow categorization of subject performance as Very Good, Good, Average, Poor or Very Poor.

- With few exceptions, the data obtained in this project showed remarkable consistency across task batteries and within task types both in terms of actual dependent measure values and general response characteristics.

- The reliability of the tasks varied by task type and dependent measure. In general, response time measures of various tasks yielded acceptable, and in some cases very good, reliability indices. Percentage correct measures of tasks were almost uniformly unacceptable, due most probably to ceiling effects. Timing tasks (WRAIR PAB Time Wall and Interval Production) varied in their reliability, but neither yielded an impressive level of reliability.

- Comparisons of similar tasks across batteries yielded a variety of findings. Significant differences were sometimes found between versions of the same task, however, these differences were often not greater in magnitude than the difference associated with the day-to-day changes experienced over training sessions. The Mathematical Processing and Memory Search tasks yielded no differences of any importance. The percentage correct measure for the Grammatical Reasoning task was higher on the CTS version compared with the STRES version, but no difference was found for response time. The CTS Spatial Processing task had faster response times than the STRES version, but demonstrated no difference in percentage correct. The only task to show substantial differences for all dependent measures was the Unstable Tracking task. The STRES version of this task provided a smaller RMS Error and fewer Edge Violations.

- Comparison of the CTS data to a previous CTS database revealed good correspondence with the exception of the Unstable Tracking task, which had been substantially changed between CTS Version 1.0 and CTS Version 2.0, which was used in this study.

- No important differences due to the influence of task order were observed in this study.

- No important differences due to the influence of battery sequence were observed in this study.

- Response deadlines provided a faster mean response time but at the expense of more missed responses when actual deadlines were imposed. However, when subject response urgency was increased through instructions, the faster mean response times were not accompanied by a significantly lower percentage correct.

- Results from the Extended Trial Length analysis revealed that during the first three-minute epoch of performance, subjects appeared to perform at about the same level regardless of the overall trial length. However, average performance across individual trial lengths, differed from one trial length to another. There was some evidence that subjects performed the continuous Unstable Tracking task more poorly and more erratically over an extended period of time. They also appeared to show improvement in response time for some discrete tasks over the extended trial length, but this may have been the result of a subtle speed-accuracy tradeoff.

- Preliminary analyses suggest that the psychometric state measures included in this study have potential for effectively assessing changes in psychological state.

REFERENCES

Acton, W.H., and Crabtree, M.S. (1985). *User's Guide for the Criterion Task Set* (Tech. Report AAMRL-TR-85-034). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

AGARD (NATO - Advisory Group for Aerospace Research & Development), Aerospace Medical Panel Working Group 12 (1989). *Human Performance Assessment Methods* (AGARDograph No. 308). Neuilly sur Seine, France.

Asch, S.E. (1951). Effects of group pressure upon the modification and distortion of judgement. In Z.H. Guetzkow (ed.), *Groups, leadership, and men*. Pittsburgh: Carnegie.

Asch, S.E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs, 70* (9,Whole No. 416).

Baird, P., Kasbaum, T., and Schlegel, R.E. (1990). *The Effects of Speed Stress on Cognitive Performance*. (unpublished report) Norman, OK: School of Industrial Engineering, University of Oklahoma.

Bittner, A.C., Jr., Carter, R.C., Kennedy, R.S., Harbeson, M.M., and Krause, M. (1984). *Performance Evaluation Tests for Environmental Research (PETER): Review of 112 Measures*. (Research Report No. NBDL-84R006). New Orleans, LA: Naval Biodynamics Laboratory.

Bond, C.F., and Titus, L.J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin, 94*, 265-292.

Broadbent, D. (1958). *Perception and communication*. Oxford: Pergamon Press.

Englund, C.E., Reeves, D.L., Shingledecker, C.A., Thorne, D.R., Wilson, K.P., and Hegge, F.W. (1985). *Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB), I: Design and Specification of the Battery* (JWGD3 MILPERF Report No. 85-1). Fort Detrick, MD: U.S. Army Research and Development Command.

Englund, C.E., Reeves, D.L., Shingledecker, C.A., Thorne, D.R., Wilson, K.P., and Hegge, F.W. (1987). *Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB), I: Design and Specification of the Battery* (Tech. Report NHRC-TR-87-10). San Diego, CA: Naval Health Research Center.

Guerin, B. (1986). Mere presence effects in humans: A review. *Journal of Personality and Social Psychology, 22*, 38-77.

Guilford, J.P. (1954). *Psychometric methods*. New York: McGraw-Hill.

Guilford, J.P. (1956). *Fundamental statistics in psychology and education*. New York: McGraw-Hill.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.

Hegge, F.W., Reeves, D.L., Poole, D.P., and Thorne, D.R. (1985). *The Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB), II: Hardware/Software Design and Specifications* (JWGD3 MILPERF Report No. 85-2). Washington, DC: Walter Reed Army Institute of Research.

Kerr, B. (1973). Processing demands during mental operations. *Memory and Cognition, 1*, 401-412.

Lambert, M.E. (1990). *The Effects of Site Configuration on a Tactile Information Display for the Human Head.* Unpublished master's thesis. Norman, OK: The University of Oklahoma.

McGrath, J.E. (1962). The influence of positive interpersonal relations on adjustment and effectiveness in rifle teams. *Journal of Abnormal and Social Psychology, 65*, 365-375.

McNair, D.M., Lorr, M., and Droppleman, L.F. (1971). *Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.

Myers, A. (1980). *Experimental psychology.* New York: van Nostrand.

Myers, A.E. (1962). Team competition, success, and the adjustment of group members. *Journal of Abnormal and Social Psychology, 65*, 325-332.

Navon, D., and Gopher, D. (1979). On the economy of the human processing system. *Psychological Review, 86*, 214-255.

Norman, D. (1968). Toward a theory of memory and attention. *Psychological Review, 75*, 522-536.

Nunnally, J.C. (1967). *Psychometric theory*. New York: McGraw-Hill.

Perez, W.A., Masline, P.J., Ramsey, E.G., and Urban, K.E. (1987). *Unified Tri-service Cognitive Performance Assessment Battery: Review and Methodology* (Tech. Report AAMRL-TR-87-007). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Reeves, D.L., Thorne, D.R., Winter, S.L., and Hegge, F.W., (1989). *The Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB), II: Hardware/Software Design and Specifications* (Tech. Report NAMRL SR89-1). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Reeves, D.L., Winter, K.P., LaCour, S.J., Winter, K.M., Vogel, K., and Grissett, J.D., (1990). *The UTCPAB-AGARD STRES Battery: User's Manual and System Documentation* (Tech. Report NAMRL SR90-1). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Sanders, A.F. (1979). Some remarks on mental load. In N. Moray (ed.), *Mental workload: its theory and measurement.* New York: Plenum.

Schlegel, B., Schlegel, R.E., and Gilliland, K. (1988). Determining Excessive Mental Workload with the Subjective Workload Assessment Technique. In *Proceedings of the Tenth Congress of the International Ergonomics Association*, Vol II, 475-477, Sydney, Australia.

Schlegel, R.E. (1986). *Development of an Optimal Testing Protocol for the USAF Criterion Task Set (CTS)* (Final Scientific Report for contract SCEEE-84 RIP 47). Norman, OK: University of Oklahoma, School of Industrial Engineering.

Schlegel, R.E., and Gilliland, K. (1986). *To Determine the Existence of a Subjective Measure of Excessive Mental Workload in a Single Cognitive Task.* Contract Report PO246442. Norman, OK: The University of Oklahoma.

Schlegel, R.E., and Gilliland, K. (1990). *Evaluation of the Criterion Task Set - Part I: Appendices A and B - Univariate Summaries* (Tech. Report AAMRL-TR-90-008). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Schlegel, R.E., and Gilliland, K. (1990). *Evaluation of the Criterion Task Set - Part I: CTS Performance and SWAT Data - Baseline Conditions* (Tech. Report AAMRL-TR-90-007). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Schlegel, R.E., and Gilliland, K. (1990). *Pilot Study of Vibro-Tactile Stimulation of the Human Head for Information Display*. Contract Report AB2011. Norman, OK: The University of Oklahoma.

Schlegel, R.E., and Shingledecker, C.A. (1985). Training Characteristics of the Criterion Task Set Workload Assessment Battery. In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 770-773). Santa Monica, CA: Human Factors Society.

Selye H. (1976). *The stress of life* (Rev. ed.). New York: McGraw-Hill.

Shingledecker, C.A. (1984). *A Task Battery for Applied Human Performance Assessment Research* (Tech. Report AFAMRL-TR-84-071). Wright-Patterson Air Force Base, OH: Air Force Aerospace Medical Research Laboratory.

Shingledecker, C.A., Acton, W.H., and Crabtree, M.S. (1983, October). *Development and Application of a Criterion Task Set for Workload Metric Evaluation* (SAE Technical Paper 831419). Aerospace Congress & Exposition, Long Beach, California.

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist, 57*, 421-457.

Thorndike, R.L. (1951). Reliability. In E.F. Lindquist (ed.), *Educational measurement*. Washington, D.C.: American Council on Education.

Thorne, D.R. (1990). *Walter Reed Performance Assessment Battery*. Washington, D.C.: Walter Reed Army Institute of Research.

Thorne, D.R., Genser, S.G., Sing, H.C., and Hegge, F.W. (1985). The Walter Reed Performance Assessment Battery. *Neurobehavioral Toxicology and Teratology, 7*, 415-418.

Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review, 76*, 282-299.

Underwood, B.J., and Shaughnessy, J.J. (1975). *Experimentation in psychology*. New York: Wiley.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (ed.), *Attention and performance VIII*. Hillsdale, NJ: Erlbaum.

Wickens, C.D. (1981). *Processing Resources in Attention, Dual Task Performance, and Workload Assessment* (Tech. Report EPL-81-3). University of Illinois, Illinois: Engineering Psychology Research Laboratory.

Wickens, C.D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: HarperCollins Publishers Inc.

Zajonc, R.B. (1965). Social facilitation. *Science, 149*, 269-274.

APPENDIX A

SUBJECT INSTRUCTIONS FOR DEADLINE TESTING

Subject Instructions for Deadline Study - First Day (all subjects)

Until today you have been asked to respond as quickly and accurately as possible but there have been very liberal time deadlines placed on your performance (typically 15 seconds).

Starting today you will train and be tested under conditions which limit the amount of time you have to respond. Initially, this limit will be indicated by the stimulus disappearing from the screen.

There are two levels of deadline time limits: MODERATE and VERY SHORT. They are different for each task. You will be informed of the level at the start of each session.

> Attempt to maintain your current level of accuracy, but
> PLEASE TRY TO RESPOND BEFORE THE STIMULUS DISAPPEARS!!

Subject Instructions for Deadline Study - Second Day (all subjects)

Today, your first session will be under the same deadlines as yesterday.

The remaining sessions will be under three different deadline levels: NO deadline, a MODERATE deadline, and a VERY SHORT deadline, but not necessarily in that order. You will be informed of the deadline level at the start of each session.

> Attempt to maintain your current level of accuracy, but
> PLEASE TRY TO RESPOND BEFORE THE STIMULUS DISAPPEARS!!

Subject Instructions for Deadline Study - Third Day (all subjects)

Today, the three sessions will be under different deadline levels as they were yesterday. When everyone has completed all three sessions, we will meet as a group to discuss the experiment.

Subject Instructions for No Deadline

The following session will be conducted under no response deadline as during all the training sessions. While attempting to maintain a high level of accuracy, please respond as quickly as possible.

Subject Instructions for Moderate Deadline - Actual Deadline

The following session will be conducted under a MODERATE response deadline. While attempting to maintain a high level of accuracy,

PLEASE RESPOND BEFORE THE STIMULUS DISAPPEARS FROM THE SCREEN!

Subject Instructions for Moderate Deadline - Pseudo Deadline

The following session will be conducted under a MODERATE response deadline. The stimulus may not disappear from the screen but the deadline cutoff is still in place. While attempting to maintain a high level of accuracy, PLEASE RESPOND WITHIN THE DEADLINE!

Subject Instructions for Short Deadline - Actual Deadline

The following session will be conducted under a VERY SHORT response deadline. While attempting to maintain a high level of accuracy,

PLEASE RESPOND BEFORE THE STIMULUS DISAPPEARS FROM THE SCREEN!

Subject Instructions for Short Deadline - Pseudo Deadline

The following session will be conducted under a VERY SHORT response deadline. The stimulus may not disappear from the screen but the deadline cutoff is still in place. While attempting to maintain a high level of accuracy, PLEASE RESPOND WITHIN THE DEADLINE!

APPENDIX B

SCHEDULES FOR EXTENDED TRIAL LENGTH TESTING

Schedule for Extended Trial Study - Group A1

MONDAY	TUESDAY	WEDNESDAY DAY 1	THURSDAY DAY 2	FRIDAY DAY 3
Re-test	Re-test	*Sleep/Mood*	*Sleep/Mood*	*Sleep/Mood*
		24-minute Sternberg	24-minute Combo	24-minute Math
		Sleep/Mood	*Sleep/Mood*	*Sleep/Mood*
		24-minute Tracking	24-minute Gram Reas	24-minute Spatial
Re-test	Re-test	*Sleep/Mood*	*Sleep/Mood*	*Sleep/Mood*
		12-minute Math	12-minute Combo	12-minute Tracking
		6-minute Combo	6-minute Gram Reas	12-minute Sternberg
		6-minute Spatial	6-minute Combo	
		6-minute Combo	6-minute Math	6-minute Tracking
		12-minute Spatial	12-minute Gram Reas	3-min Combo
				Sleep/Mood
		3-min Combo	6-minute Sternberg	*Debriefing*
		Sleep/Mood	*Sleep/Mood*	

Schedule for Extended Trial Study - Group A2

MONDAY	TUESDAY	WEDNESDAY DAY 1	THURSDAY DAY 2	FRIDAY DAY 3
Re-test	Re-test	*Sleep/Mood*	*Sleep/Mood*	*Sleep/Mood*
		24-minute Tracking	24-minute Gram Reas	24-minute Spatial
		Sleep/Mood	*Sleep/Mood*	*Sleep/Mood*
		24-minute Sternberg	24-minute Combo	24-minute Math
Re-test	Re-test	*Sleep/Mood*	*Sleep/Mood*	*Sleep/Mood*
		12-minute Spatial	12-minute Gram	12-minute Tracking
		6-minute Combo	6-minute Combo	12-minute Sternberg
		6-minute Math	6-minute Gram	
		6-minute Combo	6-minute Spat	6-minute Tracking
		12-minute Math	12-minute Combo	3-min Combo
				Sleep/Mood
		3-min Combo	6-minute Sternberg	*Debriefing*
		Sleep/Mood	*Sleep/Mood*	

Schedule for Extended Trial Study - Group B1

MONDAY	TUESDAY	WEDNESDAY DAY 1	THURSDAY DAY 2	FRIDAY DAY 3
		Sleep/Mood	*Sleep/Mood*	*Sleep/Mood*
Re-test	**Re-test**	12-minute Math	12 -minute Combo	12 -minute Tracking
		Sleep/Mood	*Sleep/Mood*	*Sleep/Mood*
		6-minute Combo	6-minute Gram Reas	12-minute Sternberg
		6-minute Spatial	6-minute Combo	6-minute Tracking
		6-minute Combo	6-minute Math	3-min Combo
		12-minute Spatial	12-minute Gram Reas	*Sleep/Mood*
		3-min Combo	6-minute Sternberg	24-minute Math
		Sleep/Mood	*Sleep/Mood*	24-minute Spatial
Re-test	**Re-test**	24-minute Sternberg	24-minute Combo	*Sleep/Mood*
		24-minute Tracking	24-minute Gram Reas	*Debriefing*
		Sleep/Mood	*Sleep/Mood*	

Schedule for Extended Trial Study - Group B2

MONDAY	TUESDAY	WEDNESDAY DAY 1	THURSDAY DAY 2	FRIDAY DAY 3
		Sleep/Mood	*Sleep/Mood*	*Sleep/Mood*
Re-test	**Re-test**	12-minute Spatial	12 -minute Gram Reas	12 -minute Tracking
		Sleep/Mood	*Sleep/Mood*	*Sleep/Mood*
		6-minute Combo	6-minute Combo	12-minute Sternberg
		6-minute Math	6-minute Gram Reas	6-minute Tracking
		6-minute Combo	6-minute Spatial	3-min Combo
		12-minute Math	12-minute Combo	*Sleep/Mood*
		3-min Combo	6-minute Sternberg	24-minute Spatial
		Sleep/Mood	*Sleep/Mood*	24-minute Math
Re-test	**Re-test**	24-minute Tracking	24-minute Gram Reas	*Sleep/Mood*
		24-minute Sternberg	24-minute Combo	*Debriefing*
		Sleep/Mood	*Sleep/Mood*	

APPENDIX C

STRES NORMATIVE DATA

UNIVARIATE SUMMARY FOR STRES GRAMMATICAL REASONING

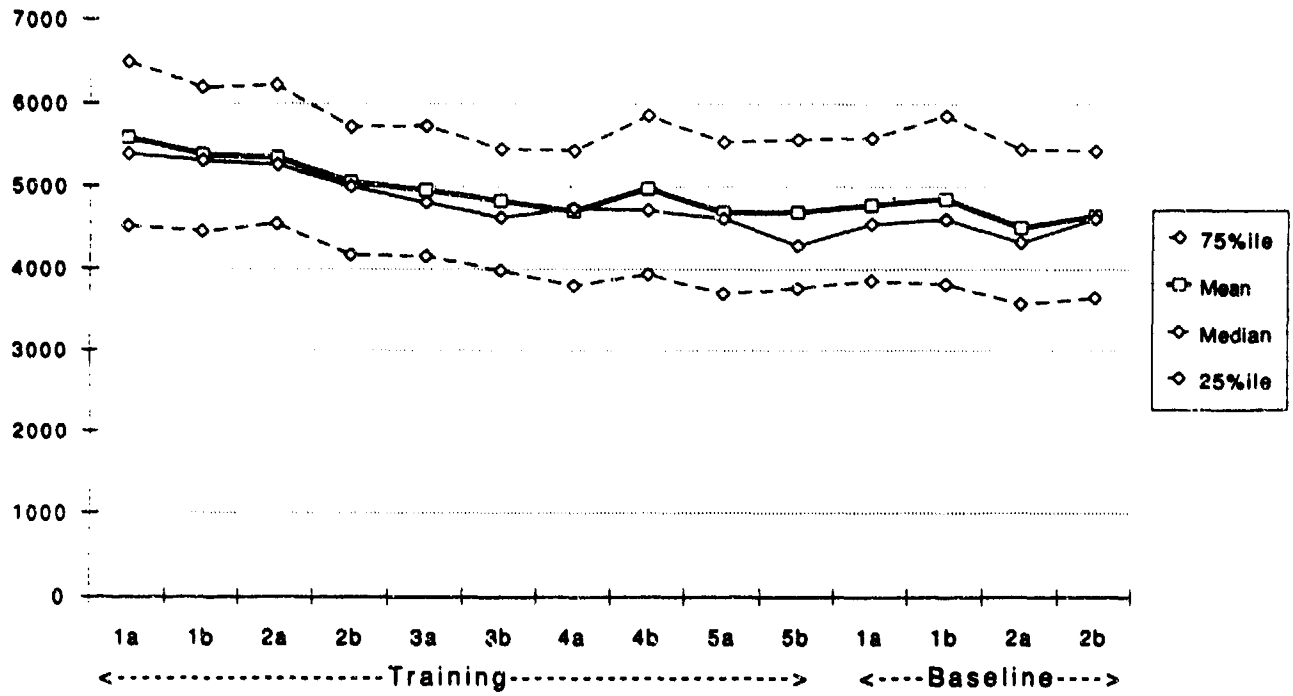
GRMMNO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
1a	5589	4122	7056	1467	5394	4527	6498
1b	5382	4015	6750	1368	5316	4463	6196
2a	5346	4161	6531	1185	5265	4555	6223
2b	5061	3801	6321	1260	5011	4180	5728
3a	4964	3788	6140	1176	4816	4162	5741
3b	4835	3649	6021	1186	4635	3989	5458
4a	4711	3558	5863	1153	4751	3804	5441
4b	4984	3458	6510	1526	4725	3942	5867
5a	4700	3415	5985	1285	4619	3705	5546
5b	4700	3351	6048	1348	4291	3772	5573
1a	4783	3418	6147	1364	4554	3861	5589
1b	4860	3363	6357	1497	4611	3819	5860
2a	4507	3232	5783	1275	4327	3579	5453
2b	4650	3280	6019	1369	4617	3656	5432

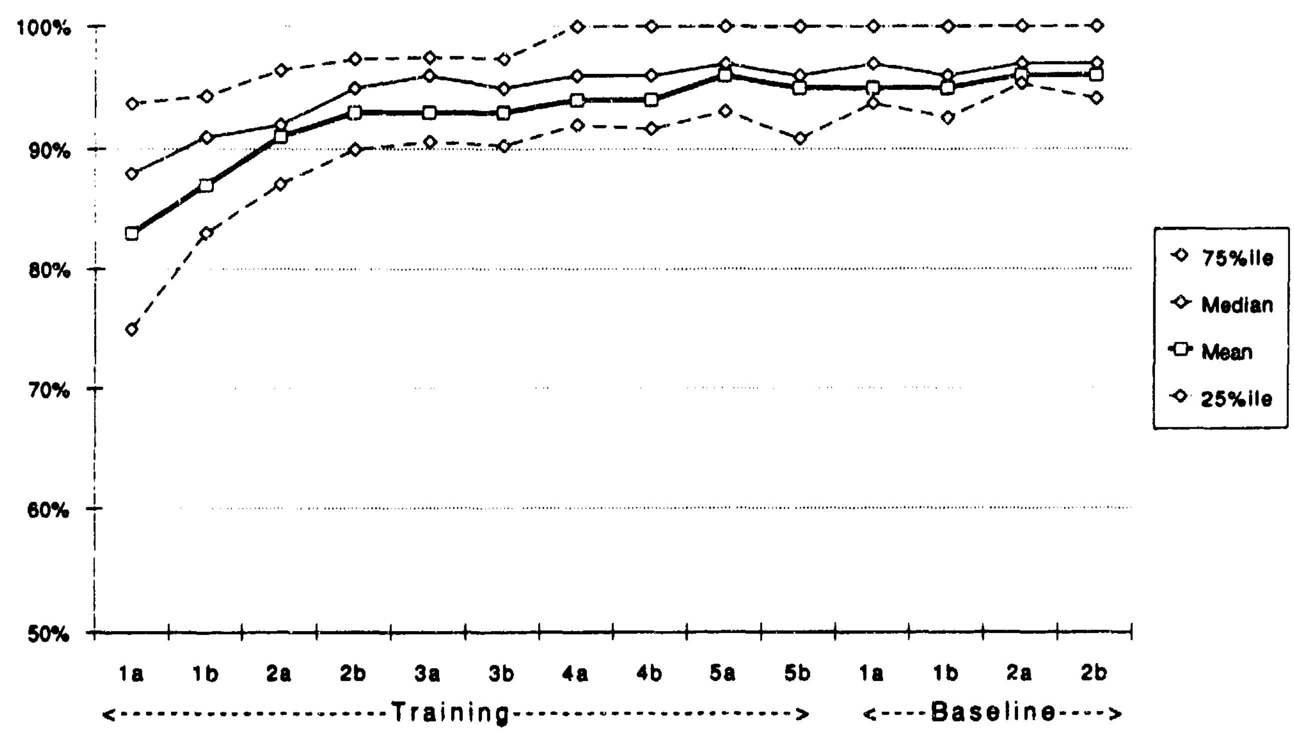
GRMPOO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
1a	83%	69%	97%	14%	88%	75%	94%
1b	87%	77%	97%	10%	91%	83%	94%
2a	91%	83%	98%	8%	92%	87%	96%
2b	93%	87%	100%	6%	95%	90%	97%
3a	93%	85%	100%	8%	96%	91%	98%
3b	93%	87%	100%	7%	95%	90%	97%
4a	94%	88%	100%	6%	96%	92%	100%
4b	94%	88%	100%	7%	96%	92%	100%
5a	96%	91%	100%	4%	97%	93%	100%
5b	95%	89%	100%	5%	96%	91%	100%
1a	95%	90%	100%	5%	97%	94%	100%
1b	95%	90%	100%	5%	96%	93%	100%
2a	96%	89%	100%	7%	97%	95%	100%
2b	96%	91%	100%	5%	97%	94%	100%

STRES Grammatical Reasoning
Mean Response Time
(msec)



STRES Grammatical Reasoning
Percentage Correct



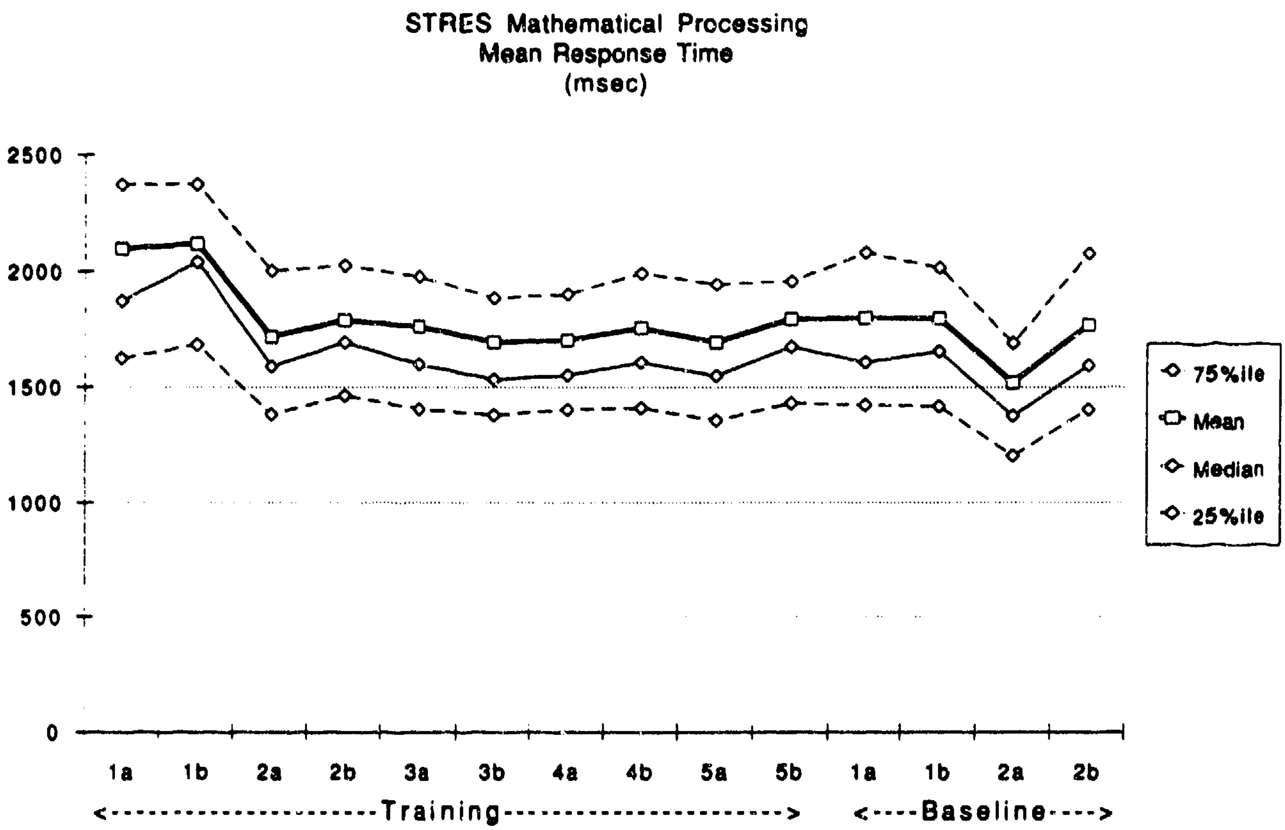
UNIVARIATE SUMMARY FOR STRES MATHEMATICAL PROCESSING

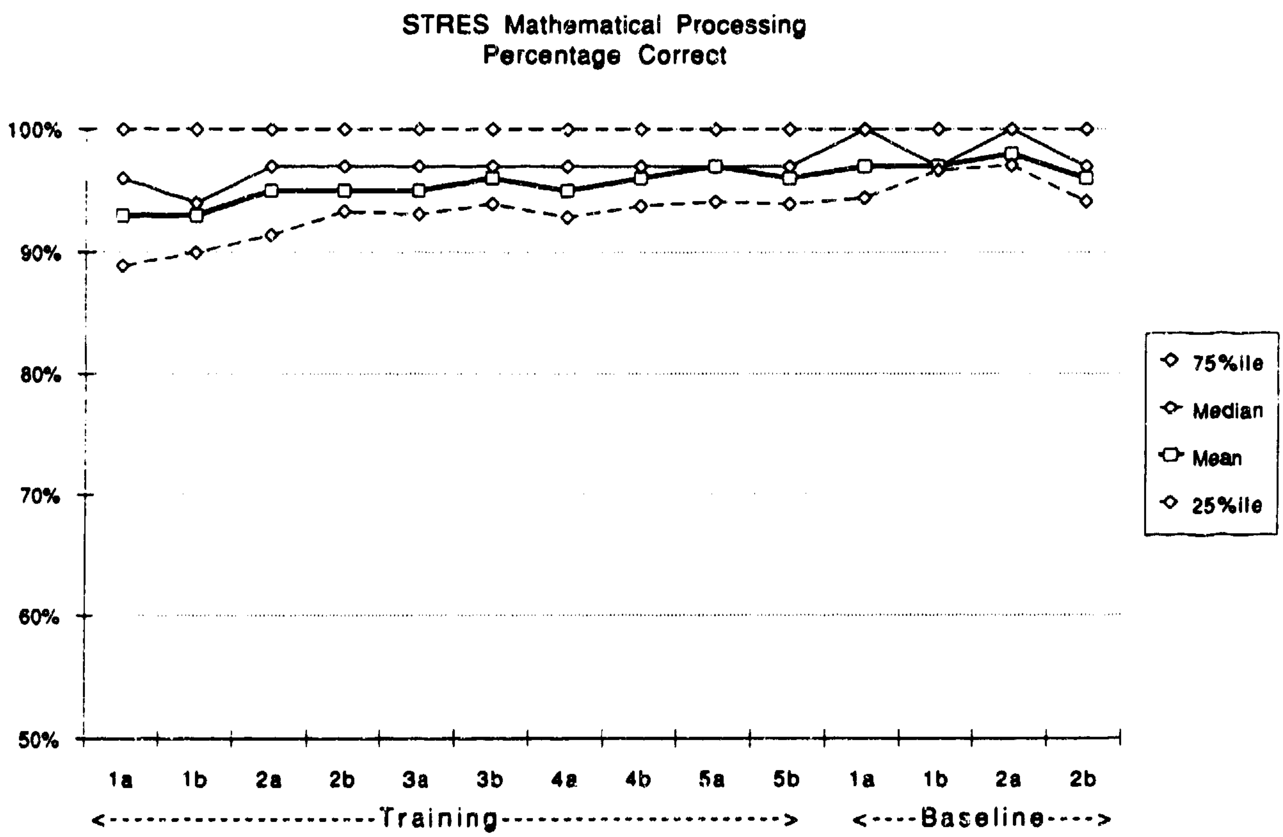
MTHMNO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
1a	2099	1415	2784	685	1873	1626	2373
1b	2123	1552	2695	571	2044	1686	2375
2a	1719	1239	2199	480	1591	1383	2004
2b	1790	1299	2281	491	1696	1466	2030
3a	1763	1201	2325	562	1602	1405	1980
3b	1695	1204	2186	491	1536	1381	1887
4a	1704	1220	2188	484	1555	1403	1903
4b	1757	1264	2249	493	1609	1412	1994
5a	1697	1242	2152	455	1553	1359	1948
5b	1797	1262	2333	535	1678	1432	1960
1a	1801	1227	2376	574	1610	1424	2084
1b	1799	1208	2390	591	1656	1418	2019
2a	1522	1082	1962	440	1378	1206	1691
2b	1770	1229	2311	541	1596	1403	2081

MTHPCO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
1a	93%	85%	100%	8%	96%	89%	100%
1b	93%	87%	99%	6%	94%	90%	100%
2a	95%	91%	100%	4%	97%	91%	100%
2b	95%	89%	100%	6%	97%	93%	100%
3a	95%	88%	100%	6%	97%	93%	100%
3b	96%	91%	100%	5%	97%	94%	100%
4a	95%	91%	100%	5%	97%	93%	100%
4b	96%	90%	100%	5%	97%	94%	100%
5a	97%	93%	100%	4%	97%	94%	100%
5b	96%	90%	100%	6%	97%	94%	100%
1a	97%	92%	100%	5%	100%	94%	100%
1b	97%	93%	100%	4%	97%	97%	100%
2a	98%	94%	100%	4%	100%	97%	100%
2b	96%	91%	100%	5%	97%	94%	100%


STRES Mathematical Processing
Mean Response Time
(msec)
2500
2000
1500
1000
500
0
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<---------------------Training----------------------> <----Baseline---->
75%ile
Mean
Median
25%ile


STRES Mathematical Processing
Percentage Correct
100%
90%
80%
70%
60%
50%
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<---------------------Training----------------------> <----Baseline---->
75%ile
Median
Mean
25%ile

UNIVARIATE SUMMARY FOR STRES MEMORY SEARCH 2

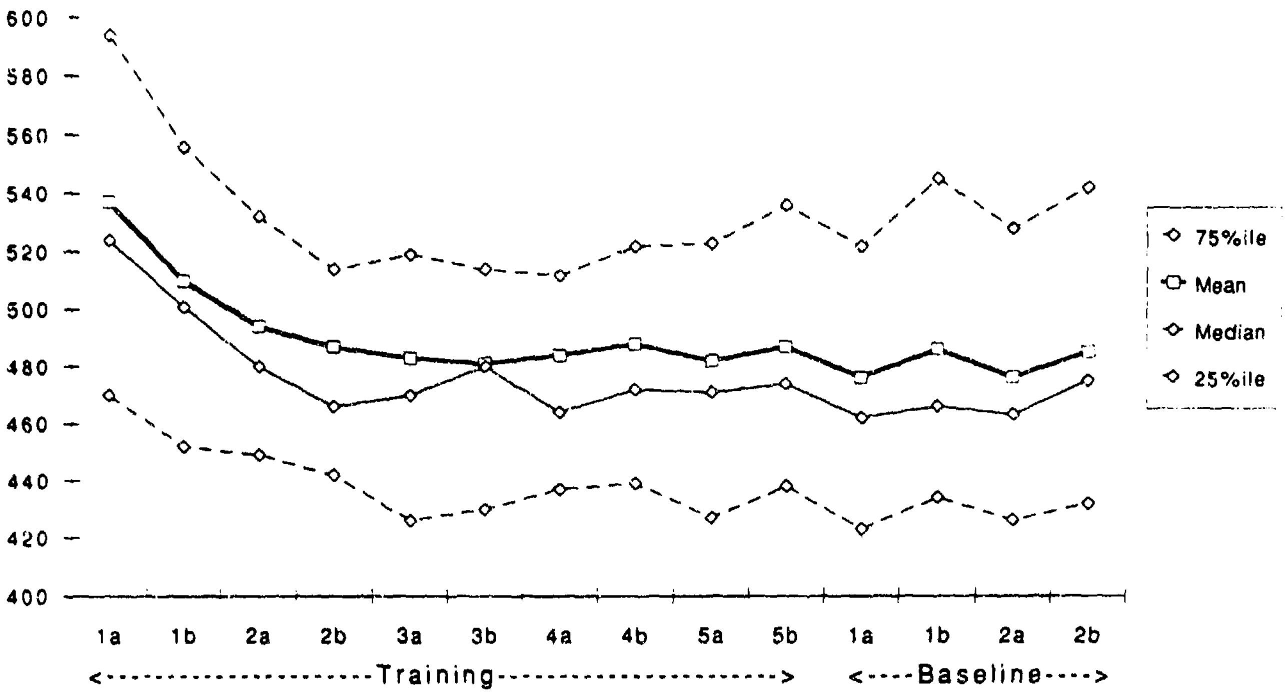
STN2MNO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
1a	537	451	624	87	524	470	594
1b	510	440	581	71	501	452	556
2a	494	424	565	71	480	449	532
2b	487	409	566	78	466	442	514
3a	483	411	555	72	470	426	519
3b	481	412	549	69	480	430	514
4a	484	409	558	74	464	437	512
4b	488	381	596	108	472	439	522
5a	482	406	557	76	471	427	523
5b	487	416	558	71	474	436	536
1a	476	409	544	67	462	423	522
1b	486	412	561	74	466	434	545
2a	476	408	545	68	463	426	528
2b	485	406	564	79	475	432	542

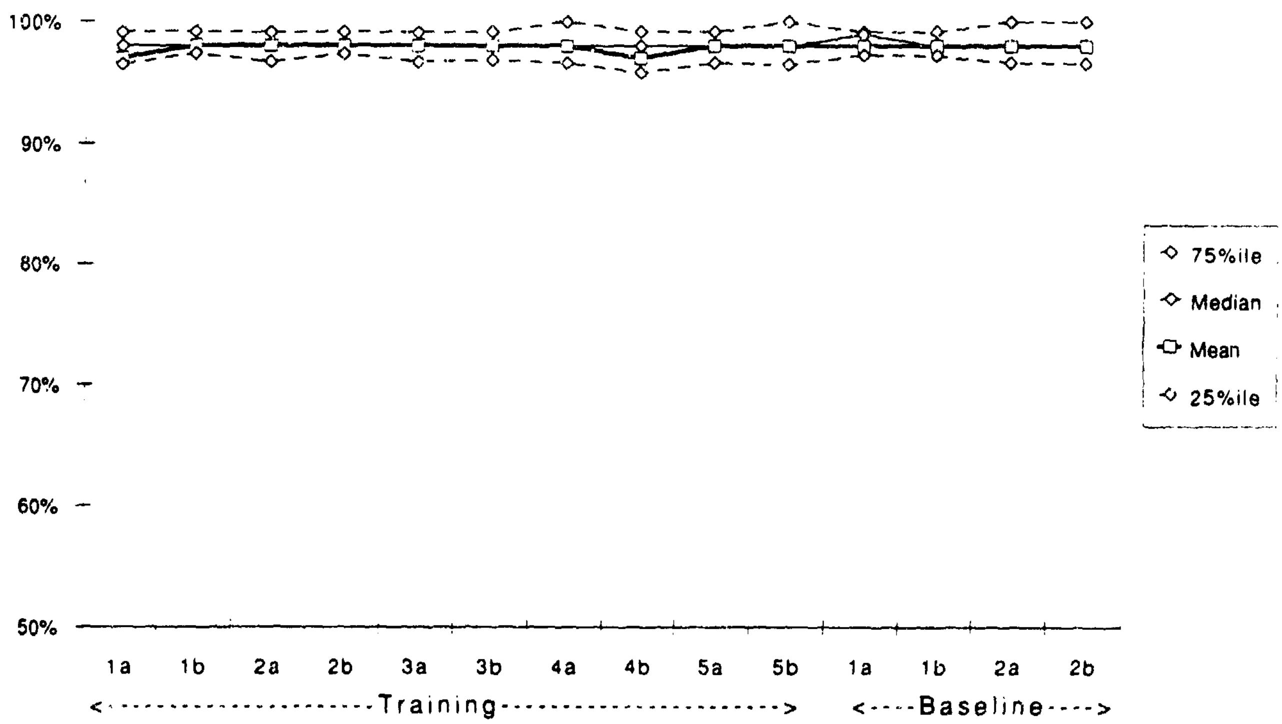
STN2PCO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
1a	97%	92%	100%	5%	98%	97%	99%
1b	98%	97%	100%	2%	98%	97%	99%
2a	98%	96%	100%	2%	98%	97%	99%
2b	98%	95%	100%	3%	98%	97%	99%
3a	98%	96%	100%	2%	98%	97%	99%
3b	98%	96%	100%	2%	98%	97%	99%
4a	98%	95%	100%	3%	98%	97%	100%
4b	97%	94%	100%	3%	98%	96%	99%
5a	98%	95%	100%	3%	98%	97%	99%
5b	98%	95%	100%	3%	98%	97%	100%
1a	98%	96%	100%	2%	99%	97%	99%
1b	98%	96%	100%	2%	98%	97%	99%
2a	98%	96%	100%	2%	98%	97%	100%
2b	98%	95%	100%	3%	98%	97%	100%

STRES Sternberg-2
Mean Response Time
(msec)



STRES Sternberg-2
Percentage Correct



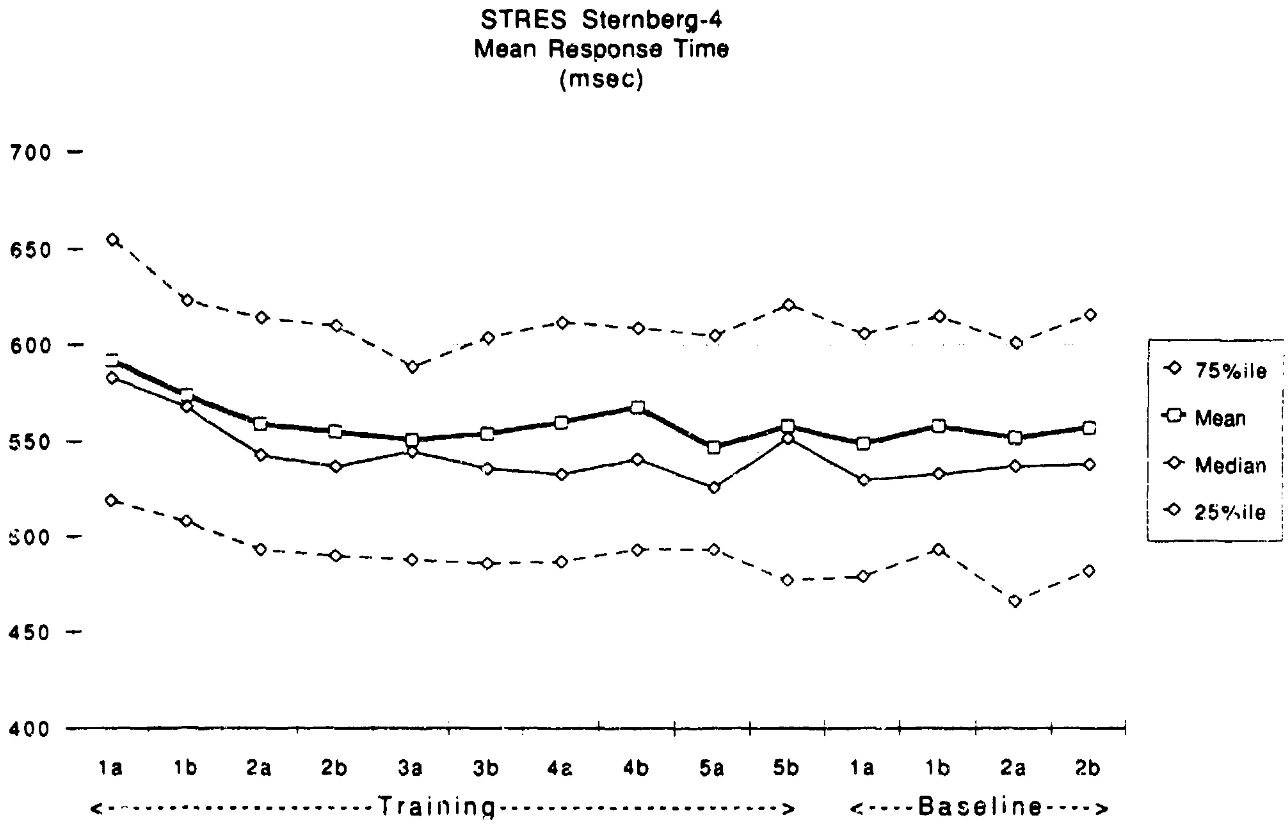
UNIVARIATE SUMMARY FOR STRES MEMORY SEARCH 4

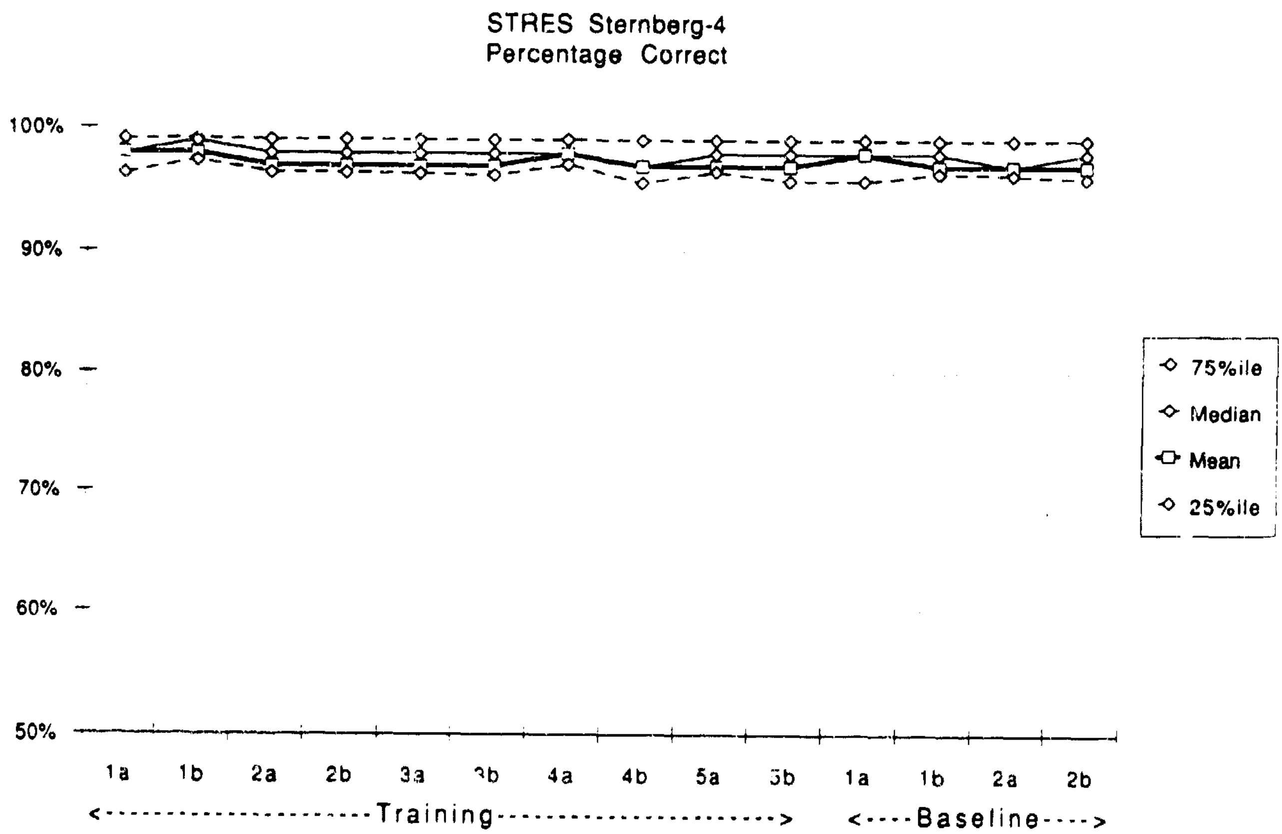
STN4MNO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
1a	592	494	691	98	583	519	655
1b	574	488	661	87	568	508	623
2a	559	477	640	82	543	493	614
2b	555	461	648	94	537	490	610
3a	551	466	637	85	545	488	589
3b	554	460	649	94	536	486	604
4a	560	438	682	122	533	487	612
4b	568	431	704	137	541	493	609
5a	547	468	625	78	526	493	605
5b	558	458	658	100	552	477	621
1a	549	455	642	93	530	479	606
1b	558	459	658	99	533	493	615
2a	552	452	651	99	537	466	601
2b	557	452	661	104	538	482	616

STN4PCO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
1a	98%	96%	100%	2%	98%	96%	99%
1b	98%	95%	100%	3%	99%	97%	99%
2a	97%	95%	100%	2%	98%	96%	99%
2b	97%	94%	100%	3%	98%	96%	99%
3a	97%	96%	99%	2%	98%	96%	99%
3b	97%	95%	100%	3%	98%	96%	99%
4a	98%	95%	100%	3%	98%	97%	99%
4b	97%	94%	100%	3%	97%	96%	99%
5a	97%	95%	100%	3%	98%	97%	99%
5b	97%	94%	100%	3%	98%	96%	99%
1a	98%	95%	100%	2%	98%	96%	99%
1b	97%	95%	100%	3%	98%	96%	99%
2a	97%	94%	100%	3%	97%	96%	99%
2b	97%	93%	100%	4%	98%	96%	99%


STRES Sternberg-4
Mean Response Time
(msec)
700
650
600
550
500
450
400
75%ile
Mean
Median
25%ile
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<---------------------Training---------------------->
<----Baseline---->


STRES Sternberg-4
Percentage Correct
100%
90%
80%
70%
60%
50%
75%ile
Median
Mean
25%ile
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<---------------------Training---------------------->
<----Baseline---->

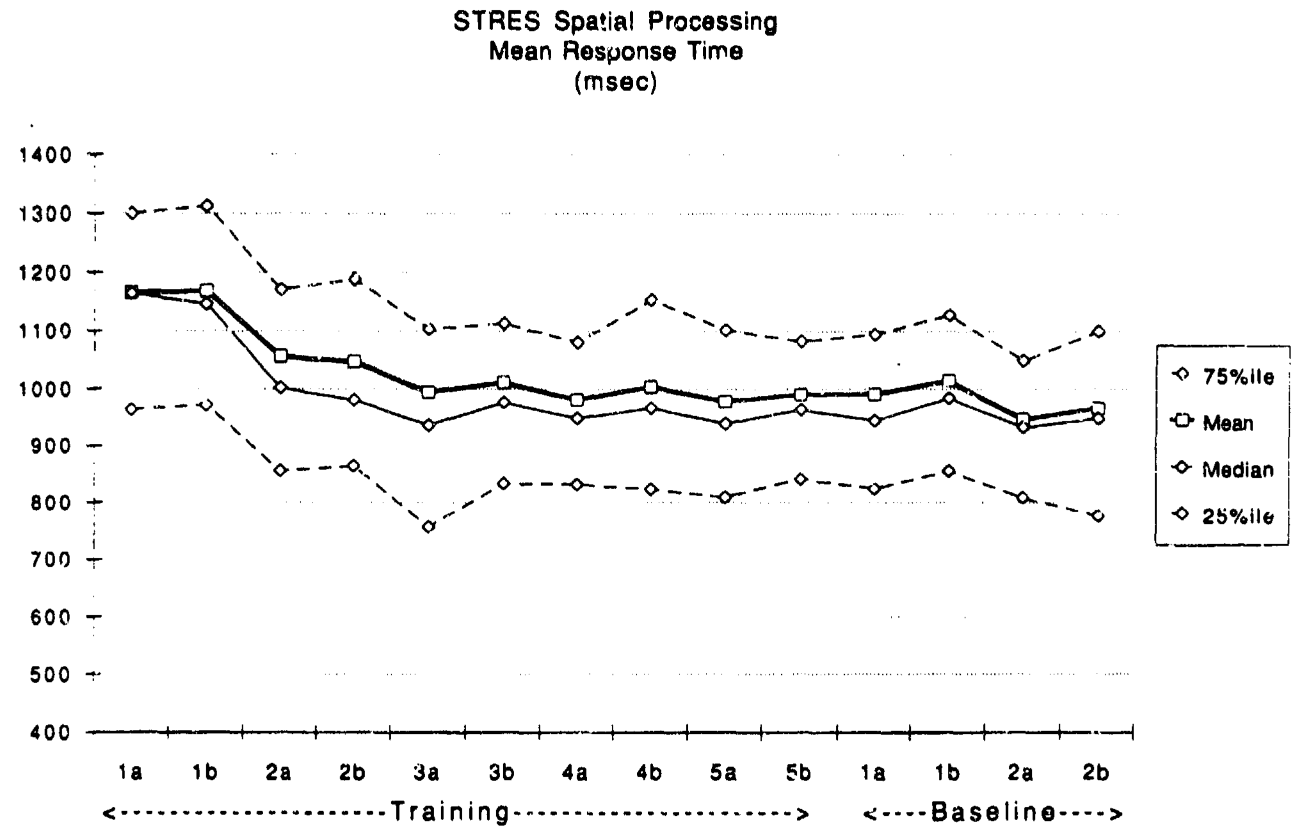
UNIVARIATE SUMMARY FOR STRES SPATIAL PROCESSING

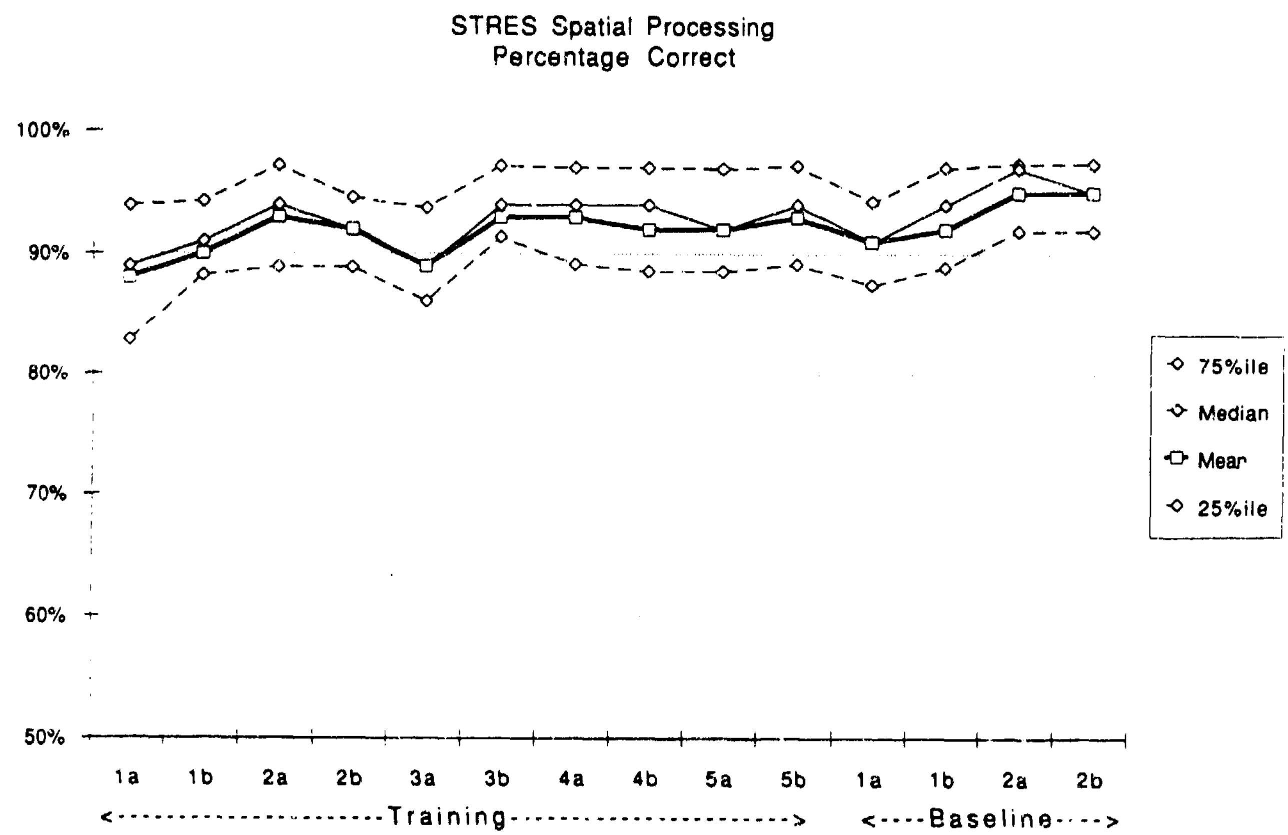
SPAMNO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
1a	1166	874	1459	292	1165	964	1301
1b	1169	874	1464	295	1147	972	1313
2a	1057	775	1338	282	1003	857	1171
2b	1048	737	1359	311	981	865	1189
3a	994	707	1281	287	937	757	1104
3b	1012	733	1291	279	977	834	1114
4a	981	697	1265	284	949	832	1081
4b	1004	738	1271	267	967	824	1154
5a	979	697	1261	282	940	810	1102
5b	991	688	1294	303	965	842	1083
1a	991	731	1251	260	945	825	1095
1b	1015	738	1292	277	985	856	1128
2a	947	716	1178	231	933	808	1050
2b	966	709	1222	257	948	776	1100

SPAPCO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
1a	88%	80%	95%	8%	89%	83%	94%
1b	90%	84%	97%	6%	91%	88%	94%
2a	93%	87%	99%	6%	94%	89%	97%
2b	92%	87%	97%	5%	92%	89%	95%
3a	89%	85%	94%	5%	89%	86%	94%
3b	93%	89%	98%	5%	94%	91%	97%
4a	93%	88%	98%	5%	94%	89%	97%
4b	92%	85%	98%	6%	94%	89%	97%
5a	92%	87%	97%	5%	92%	89%	97%
5b	93%	87%	99%	6%	94%	89%	97%
1a	91%	87%	96%	5%	91%	88%	94%
1b	92%	87%	97%	5%	94%	89%	97%
2a	95%	91%	99%	4%	97%	92%	97%
2b	95%	90%	99%	4%	95%	92%	97%


STRES Spatial Processing
Mean Response Time
(msec)
1400
1300
1200
1100
1000
900
800
700
600
500
400
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<----------------------Training---------------------->
<----Baseline---->
75%ile
Mean
Median
25%ile


STRES Spatial Processing
Percentage Correct
100%
90%
80%
70%
60%
50%
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<----------------------Training---------------------->
<----Baseline---->
75%ile
Median
Mean
25%ile

UNIVARIATE SUMMARY FOR STRES UNSTABLE TRACKING

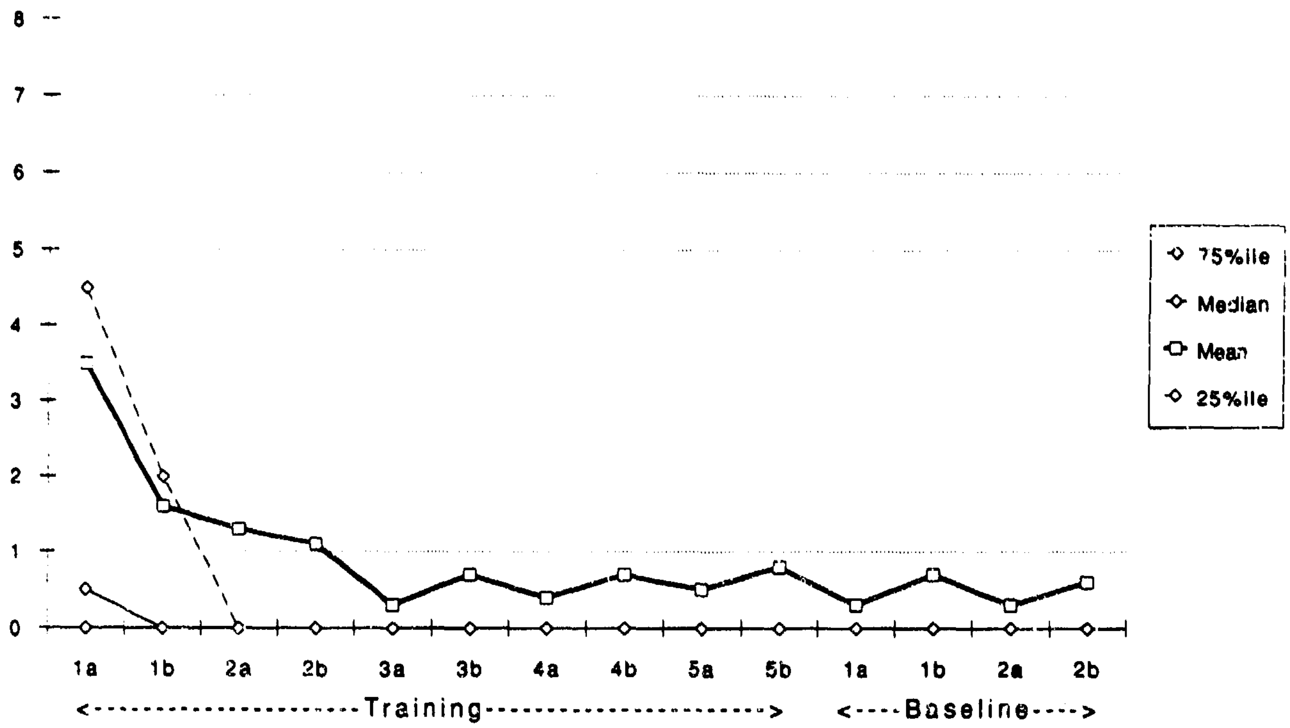
TRKEV

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
1a	3.5	0.0	9.7	6.2	0.5	0.0	4.5
1b	1.6	0.0	5.5	4.0	0.0	0.0	2.0
2a	1.3	0.0	5.9	4.6	0.0	0.0	0.0
2b	1.1	0.0	4.8	3.7	0.0	0.0	0.0
3a	0.3	0.0	1.2	0.9	0.0	0.0	0.0
3b	0.7	0.0	2.5	1.8	0.0	0.0	0.0
4a	0.4	0.0	1.7	1.4	0.0	0.0	0.0
4b	0.7	0.0	3.6	2.9	0.0	0.0	0.0
5a	0.5	0.0	2.6	2.0	0.0	0.0	0.0
5b	0.8	0.0	2.7	1.9	0.0	0.0	0.0
1a	0.3	0.0	1.4	1.1	0.0	0.0	0.0
1b	0.7	0.0	3.5	2.8	0.0	0.0	0.0
2a	0.3	0.0	1.1	0.8	0.0	0.0	0.0
2b	0.6	0.0	2.9	2.3	0.0	0.0	0.0

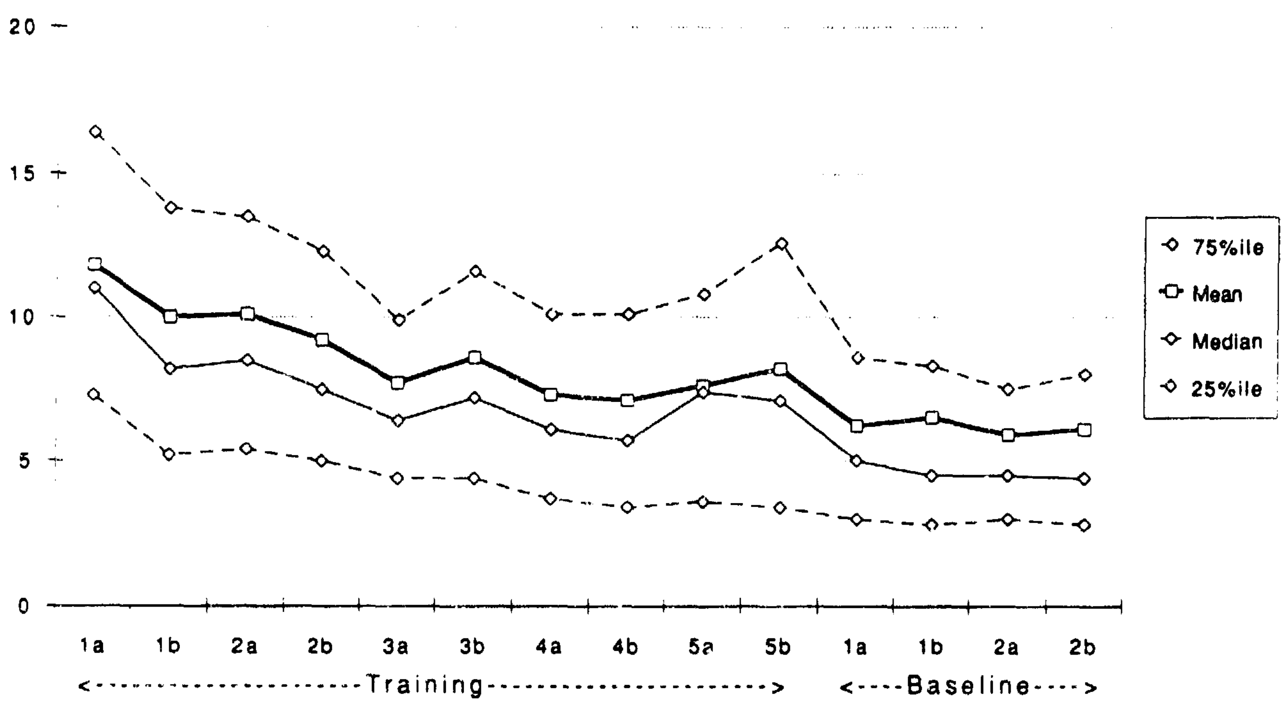
TRKRMS

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
1a	11.8	6.2	17.5	5.6	11.0	7.3	16.4
1b	10.0	3.9	16.0	6.0	8.2	5.2	13.8
2a	10.1	4.5	15.7	5.6	8.5	5.4	13.5
2b	9.2	3.5	15.0	5.7	7.5	5.0	12.3
3a	7.7	3.0	12.5	4.7	6.4	4.4	9.9
3b	8.6	3.3	13.9	5.3	7.2	4.4	11.6
4a	7.3	3.0	11.7	4.4	6.1	3.7	10.1
4b	7.1	2.5	11.7	4.6	5.7	3.4	10.1
5a	7.6	3.0	12.1	4.5	7.4	3.6	10.8
5b	8.2	2.9	13.5	5.3	7.1	3.4	12.6
1a	6.2	1.8	10.5	4.3	5.0	3.0	8.6
1b	6.5	1.3	11.8	5.3	4.5	2.8	8.3
2a	5.9	1.9	10.0	4.1	4.5	3.0	7.5
2b	6.1	1.6	10.5	4.5	4.4	2.8	8.0

STRES Unstable Tracking
Edge Violations



STRES Unstable Tracking
RMS Error



UNIVARIATE SUMMARY FOR STRES REACT TASK

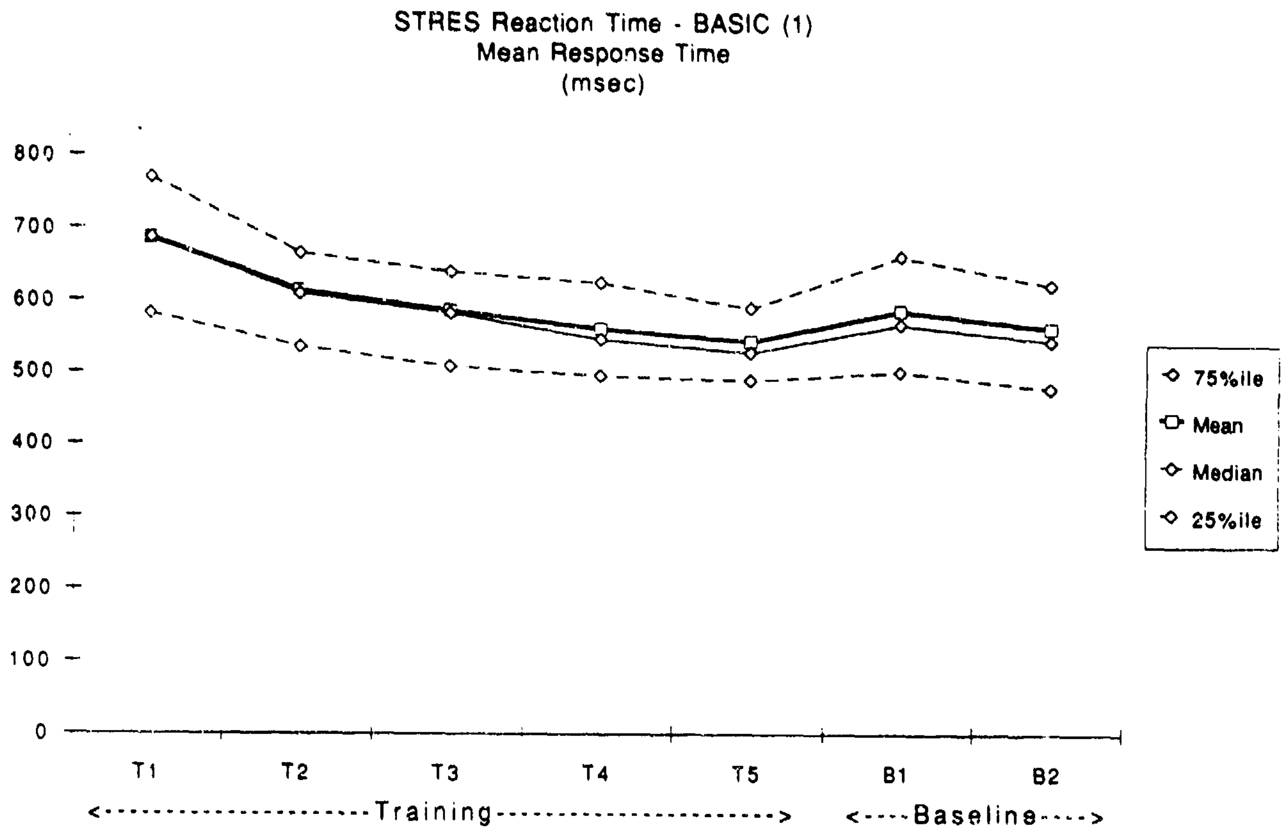
BLOCK=BASIC CODE=1

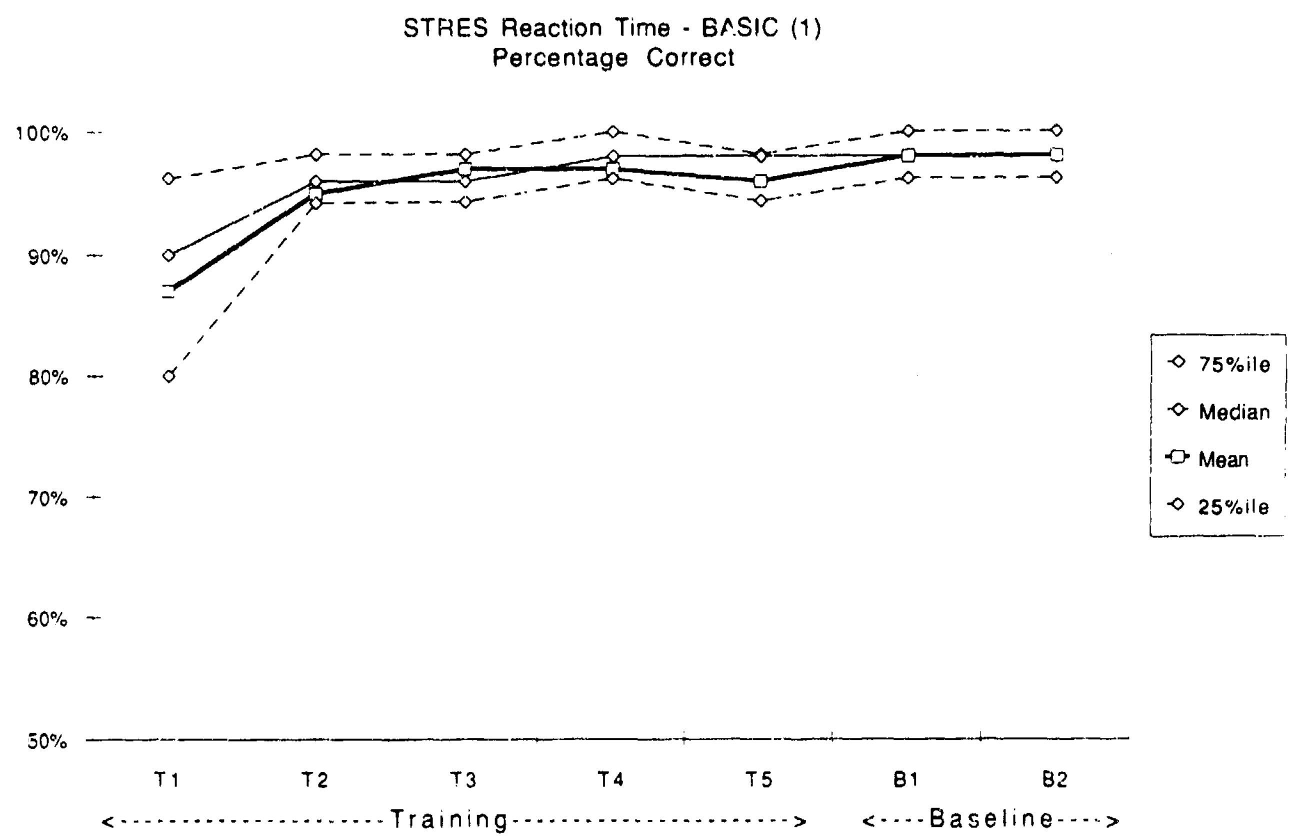
RTMN

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	687	568	805	119	687	582	770
T2	614	509	719	105	610	536	666
T3	586	494	679	92	583	509	640
T4	561	479	642	81	547	496	625
T5	544	465	623	79	529	490	590
B1	586	466	685	99	568	502	662
B2	562	468	655	93	544	478	622
21	593	493	693	100	578	509	663
22	582	479	685	103	564	507	654
26	570	470	670	100	554	511	647
27	571	461	682	110	556	485	655

RTPC

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	87%	74%	99%	13%	90%	80%	96%
T2	95%	90%	100%	5%	96%	94%	98%
T3	97%	94%	99%	3%	96%	94%	98%
T4	97%	94%	100%	3%	98%	96%	100%
T5	96%	93%	100%	4%	98%	94%	98%
B1	98%	95%	100%	3%	98%	96%	100%
B2	98%	95%	100%	3%	98%	96%	100%
21	96%	92%	100%	4%	98%	93%	98%
22	98%	95%	100%	3%	98%	96%	100%
26	97%	93%	100%	4%	98%	94%	100%
27	97%	94%	100%	3%	98%	96%	100%


STRES Reaction Time - BASIC (1)
Mean Response Time
(msec)
800
700
600
500
400
300
200
100
0
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------->
<----Baseline---->
75%ile
Mean
Median
25%ile


STRES Reaction Time - BASIC (1)
Percentage Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------->
<----Baseline---->
75%ile
Median
Mean
25%ile

UNIVARIATE SUMMARY FOR STRES REACT TASK

BLOCK=CODED CODE=2

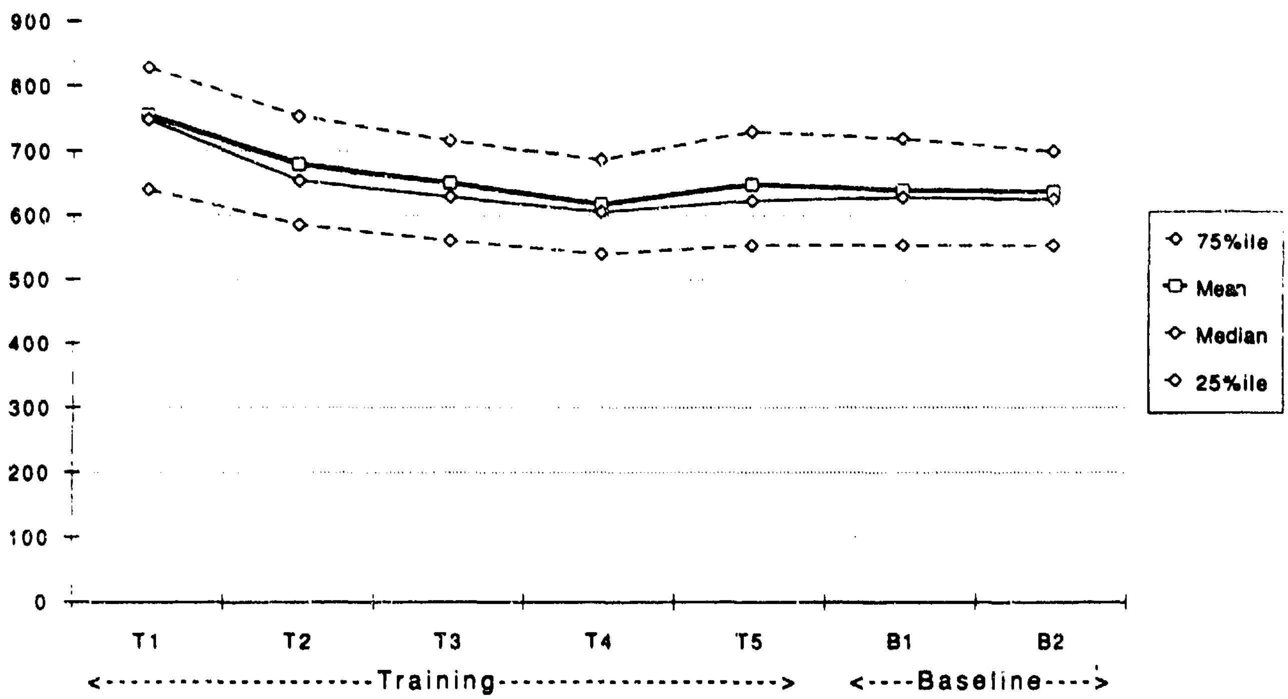
RTMN

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	756	630	881	126	749	641	829
T2	680	556	803	123	655	586	754
T3	651	535	768	116	630	562	717
T4	618	522	714	96	606	541	688
T5	649	536	763	113	624	554	731
B1	640	534	745	105	629	554	720
B2	638	526	749	112	626	554	701
21	663	522	804	141	617	563	713
22	648	528	768	120	623	557	741
26	672	546	798	126	643	563	755
27	634	504	763	130	613	533	695

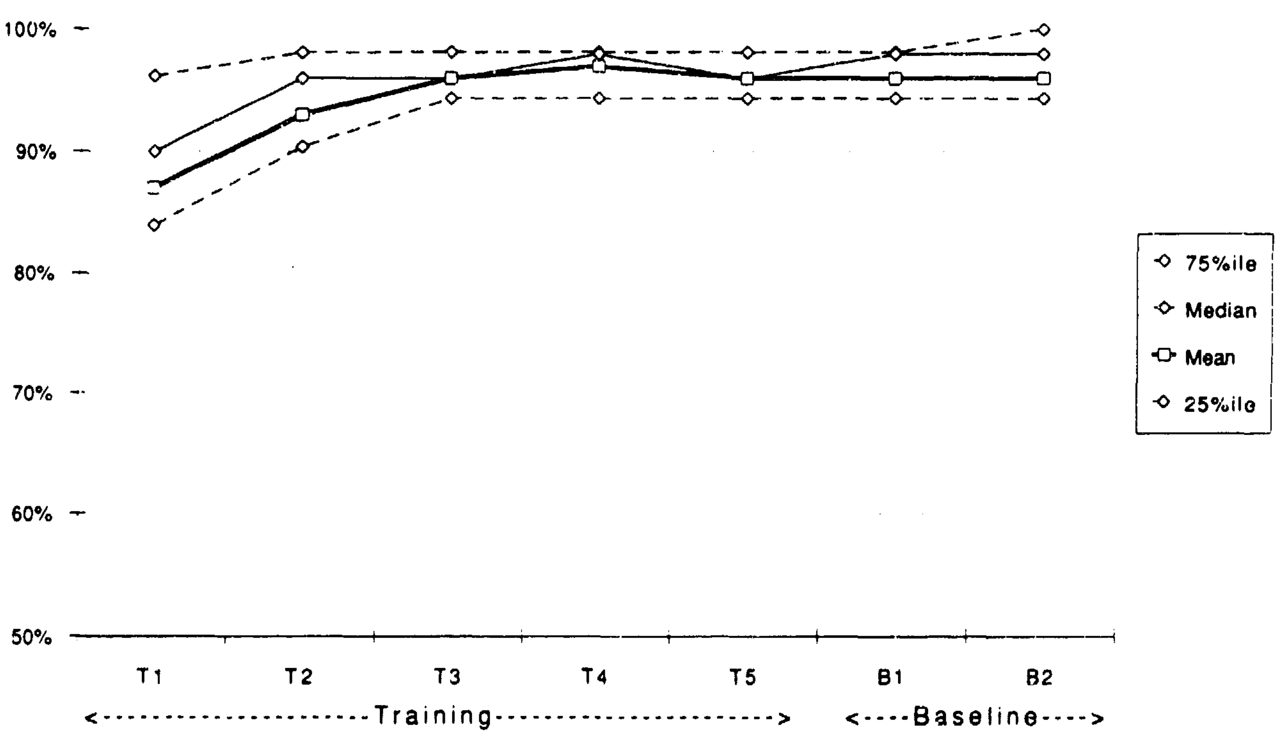
RTPC

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	87%	76%	98%	11%	90%	84%	96%
T2	93%	86%	100%	7%	96%	90%	98%
T3	96%	91%	100%	5%	96%	94%	98%
T4	97%	93%	100%	4%	98%	94%	98%
T5	96%	91%	100%	5%	96%	94%	98%
B1	96%	93%	100%	4%	98%	94%	98%
B2	96%	92%	100%	4%	98%	94%	100%
21	97%	94%	100%	3%	98%	94%	100%
22	97%	93%	100%	4%	98%	96%	100%
26	95%	89%	100%	6%	96%	92%	98%
27	96%	90%	100%	6%	98%	94%	100%

STRES Reaction Time - CODED (2)
Mean Response Time
(msec)



STRES Reaction Time - CODED (2)
Percentage Correct



UNIVARIATE SUMMARY FOR STRES REACT TASK

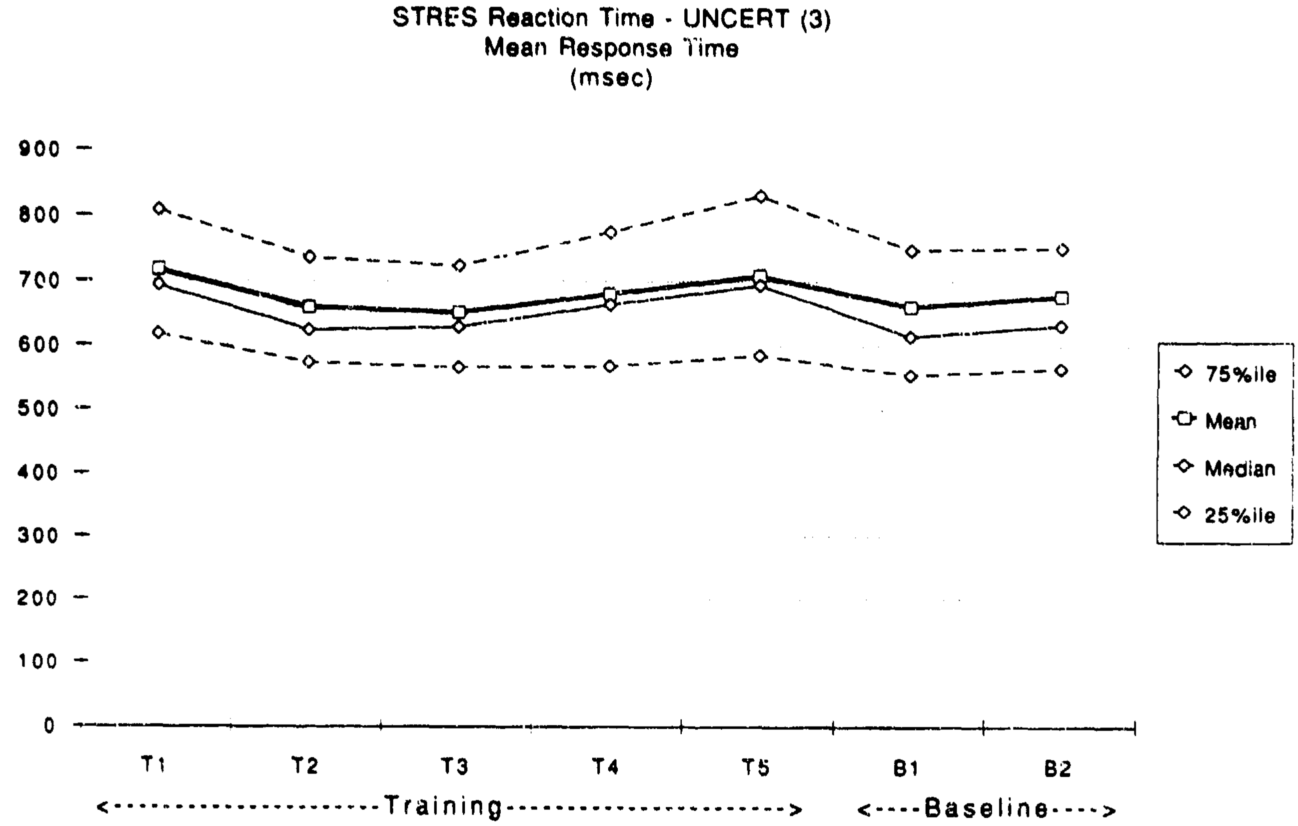
BLOCK=UNCERT CODE=3

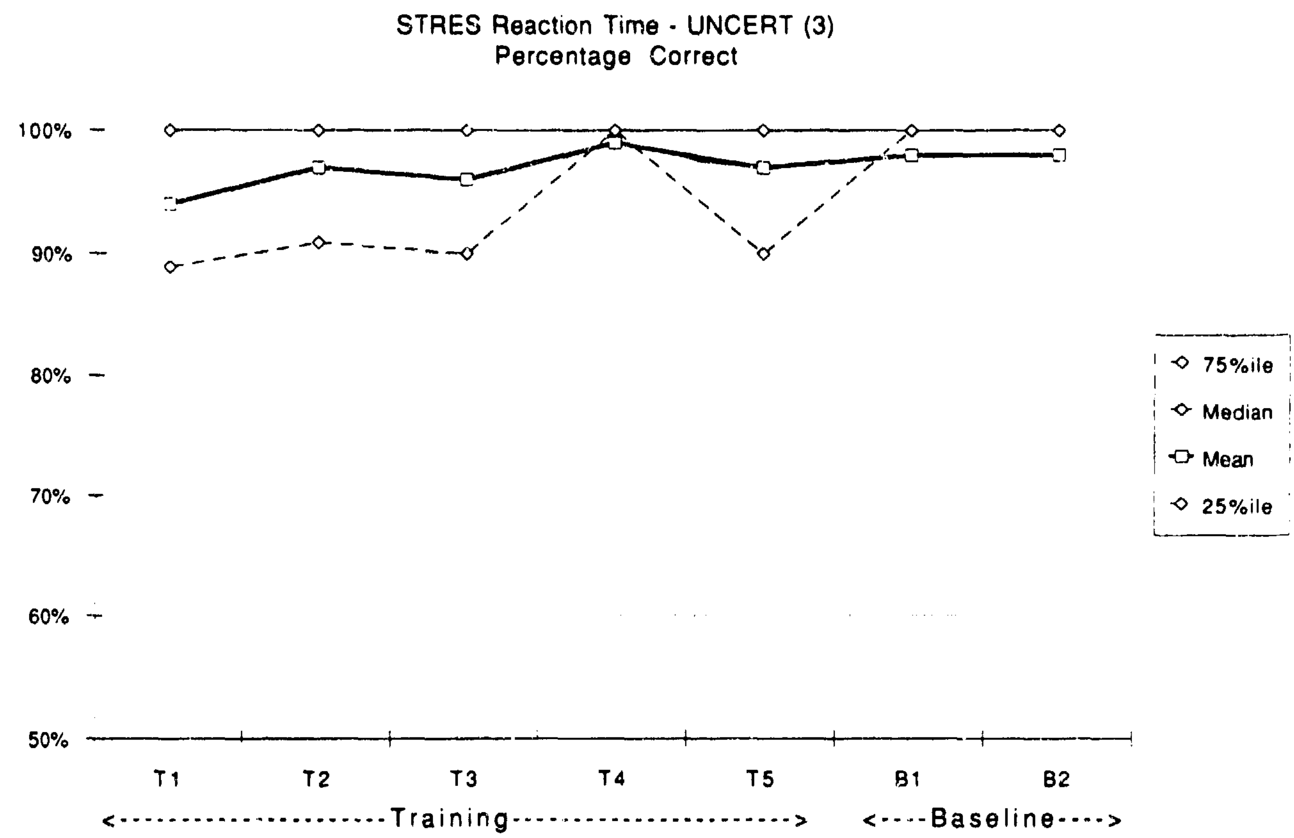
RTMN

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	718	573	862	144	694	618	809
T2	660	526	794	134	625	574	737
T3	652	543	761	109	630	567	725
T4	681	542	820	139	665	569	776
T5	709	549	869	160	695	586	831
B1	662	521	803	141	616	556	749
B2	678	520	836	158	633	565	752
21	742	500	984	242	664	564	839
22	670	518	822	152	644	552	771
26	735	561	909	174	711	604	830
27	732	564	901	168	711	591	832

RTPC

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	94%	84%	100%	10%	100%	89%	100%
T2	97%	90%	100%	6%	100%	91%	100%
T3	96%	89%	100%	7%	100%	90%	100%
T4	99%	95%	100%	4%	100%	100%	100%
T5	97%	91%	100%	5%	100%	90%	100%
B1	98%	93%	100%	5%	100%	100%	100%
B2	98%	93%	100%	5%	100%	100%	100%
21	95%	86%	100%	9%	100%	86%	100%
22	96%	90%	100%	6%	100%	89%	100%
26	98%	94%	100%	4%	100%	100%	100%
27	96%	89%	100%	7%	100%	100%	100%


STRES Reaction Time - UNCERT (3)
Mean Response Time
(msec)
900
800
700
600
500
400
300
200
100
0
T1
T2
T3
T4
T5
B1
B2
<--------------------Training-------------------->
<----Baseline---->
75%ile
Mean
Median
25%ile


STRES Reaction Time - UNCERT (3)
Percentage Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<--------------------Training-------------------->
<----Baseline---->
75%ile
Median
Mean
25%ile

UNIVARIATE SUMMARY FOR STRES REACT TASK

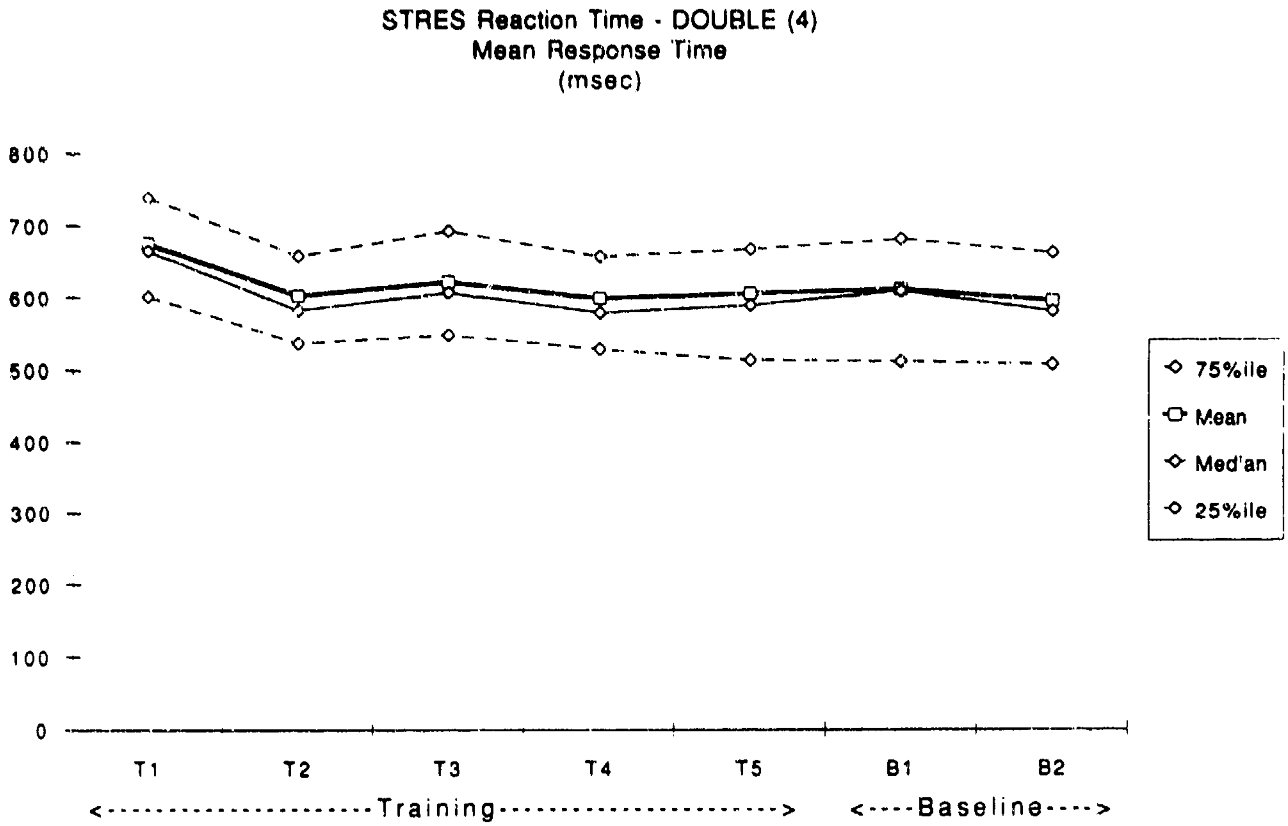
BLOCK=DOUBLE CODE=4

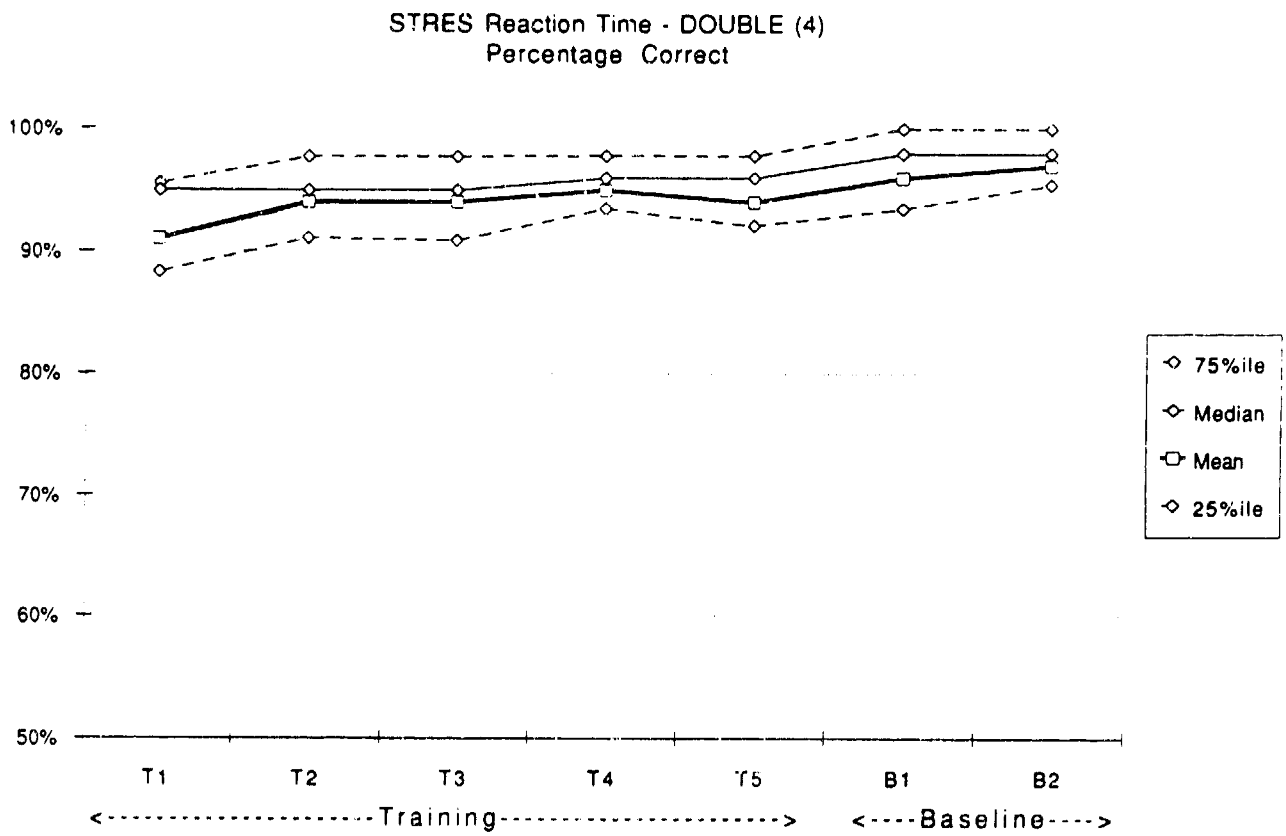
RTMN

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	675	576	774	99	665	601	739
T2	603	516	689	86	582	537	659
T3	622	522	723	101	607	543	693
T4	599	512	687	88	579	529	657
T5	605	499	710	105	589	514	667
B1	611	503	718	108	609	512	681
B2	595	491	699	104	580	507	662
21	616	517	715	99	607	525	689
22	628	512	745	116	618	536	688
26	642	515	769	127	599	547	736
27	621	500	743	121	601	517	688

RTPC

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	91%	83%	100%	9%	95%	88%	96%
T2	94%	89%	99%	5%	95%	91%	98%
T3	94%	89%	99%	5%	95%	91%	98%
T4	95%	88%	100%	7%	96%	93%	98%
T5	94%	89%	100%	5%	96%	92%	98%
B1	96%	90%	100%	6%	98%	93%	100%
B2	97%	93%	100%	4%	98%	95%	100%
21	95%	91%	100%	5%	96%	94%	98%
22	96%	91%	100%	4%	97%	93%	98%
26	96%	91%	100%	4%	96%	93%	100%
27	96%	91%	100%	5%	98%	94%	100%


STRES Reaction Time - DOUBLE (4)
Mean Response Time
(msec)
800
700
600
500
400
300
200
100
0
75%ile
Mean
Med'an
25%ile
T1
T2
T3
T4
T5
B1
B2
<----------Training---------->
<----Baseline---->


STRES Reaction Time - DOUBLE (4)
Percentage Correct
100%
90%
80%
70%
60%
50%
75%ile
Median
Mean
25%ile
T1
T2
T3
T4
T5
B1
B2
<----------Training---------->
<----Baseline---->

UNIVARIATE SUMMARY FOR STRES REACT TASK

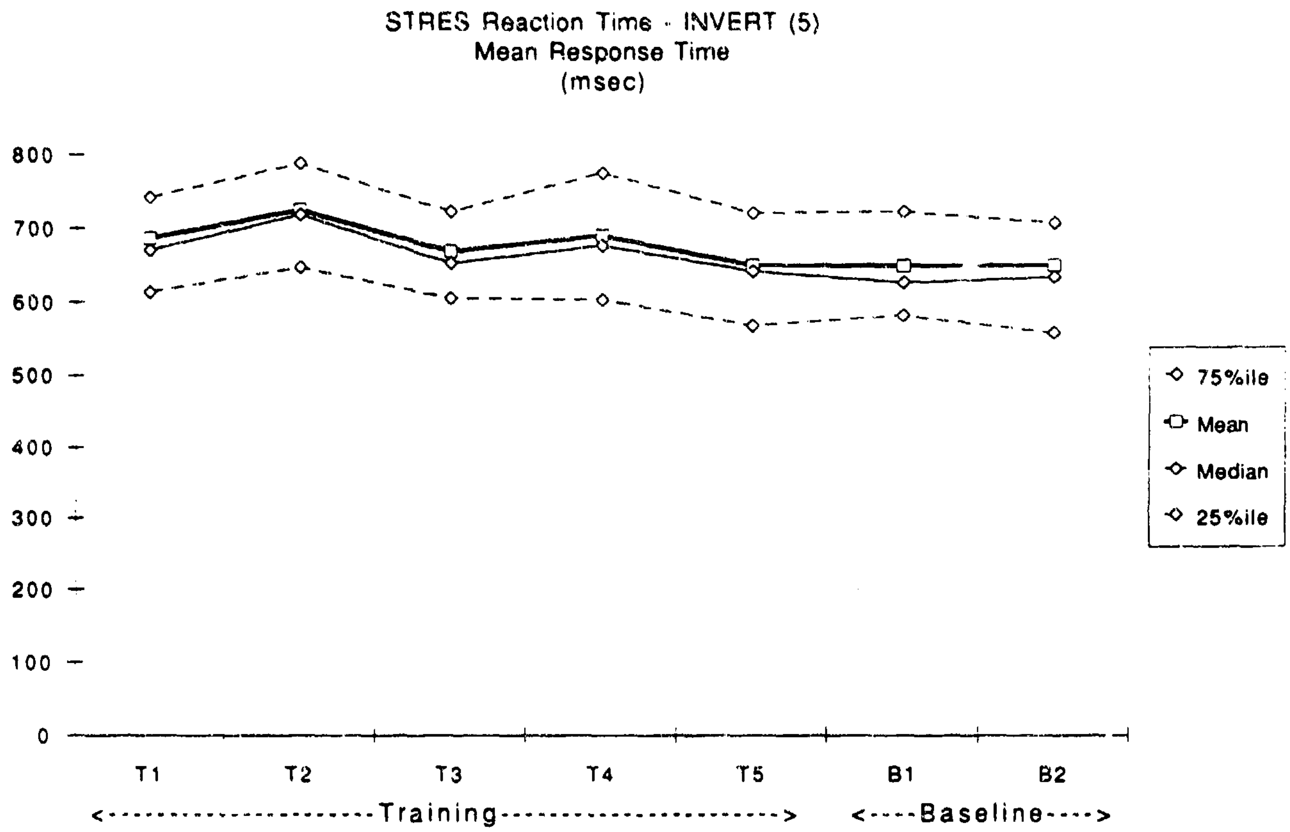
BLOCK=INVERT CODE=5

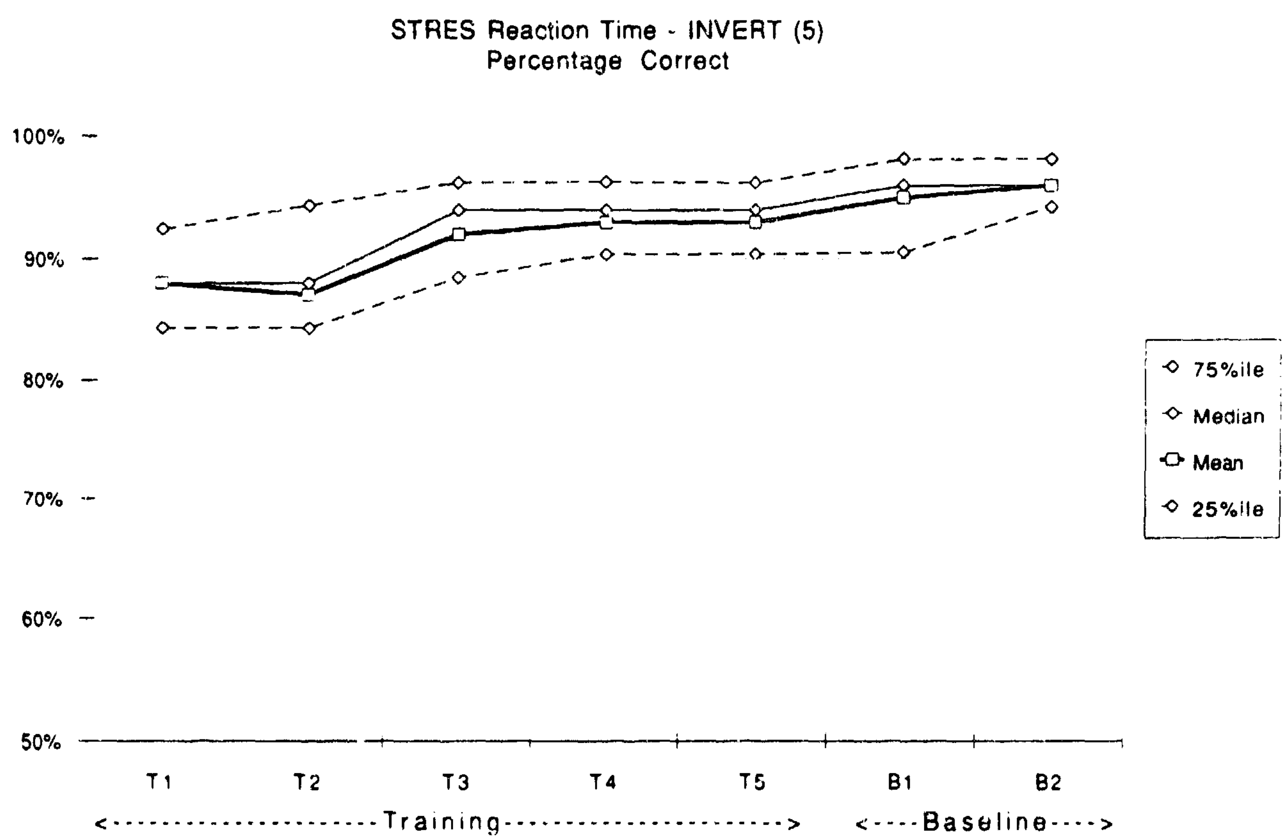
RTMN

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	687	591	783	96	671	614	742
T2	725	615	835	110	719	648	789
T3	670	571	770	100	654	606	723
T4	691	576	806	115	678	604	776
T5	651	545	757	106	643	569	721
B1	650	542	758	108	628	583	723
B2	651	531	772	120	635	559	708
21	702	553	845	143	698	590	741
22	672	540	805	133	662	591	748
26	688	553	823	135	683	602	770
27	681	547	814	134	646	552	791

RTPC

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	88%	80%	96%	8%	88%	84%	92%
T2	87%	78%	97%	9%	88%	84%	94%
T3	92%	86%	97%	6%	94%	88%	96%
T4	93%	87%	98%	6%	94%	90%	96%
T5	93%	89%	98%	5%	94%	90%	96%
B1	95%	90%	99%	5%	96%	91%	98%
B2	96%	92%	100%	4%	96%	94%	98%
21	96%	90%	100%	5%	98%	94%	99%
22	94%	90%	99%	4%	94%	91%	98%
26	96%	92%	99%	4%	96%	94%	98%
27	95%	90%	100%	5%	96%	92%	98%


STRES Reaction Time - INVERT (5)
Mean Response Time
(msec)
800
700
600
500
400
300
200
100
0
75%ile
Mean
Median
25%ile
T1
T2
T3
T4
T5
B1
B2
<--------------------Training--------------------->
<----Baseline---->


STRES Reaction Time - INVERT (5)
Percentage Correct
100%
90%
80%
70%
60%
50%
75%ile
Median
Mean
25%ile
T1
T2
T3
T4
T5
B1
B2
<--------------------Training--------------------->
<----Baseline---->

UNIVARIATE SUMMARY FOR STRES REACT TASK

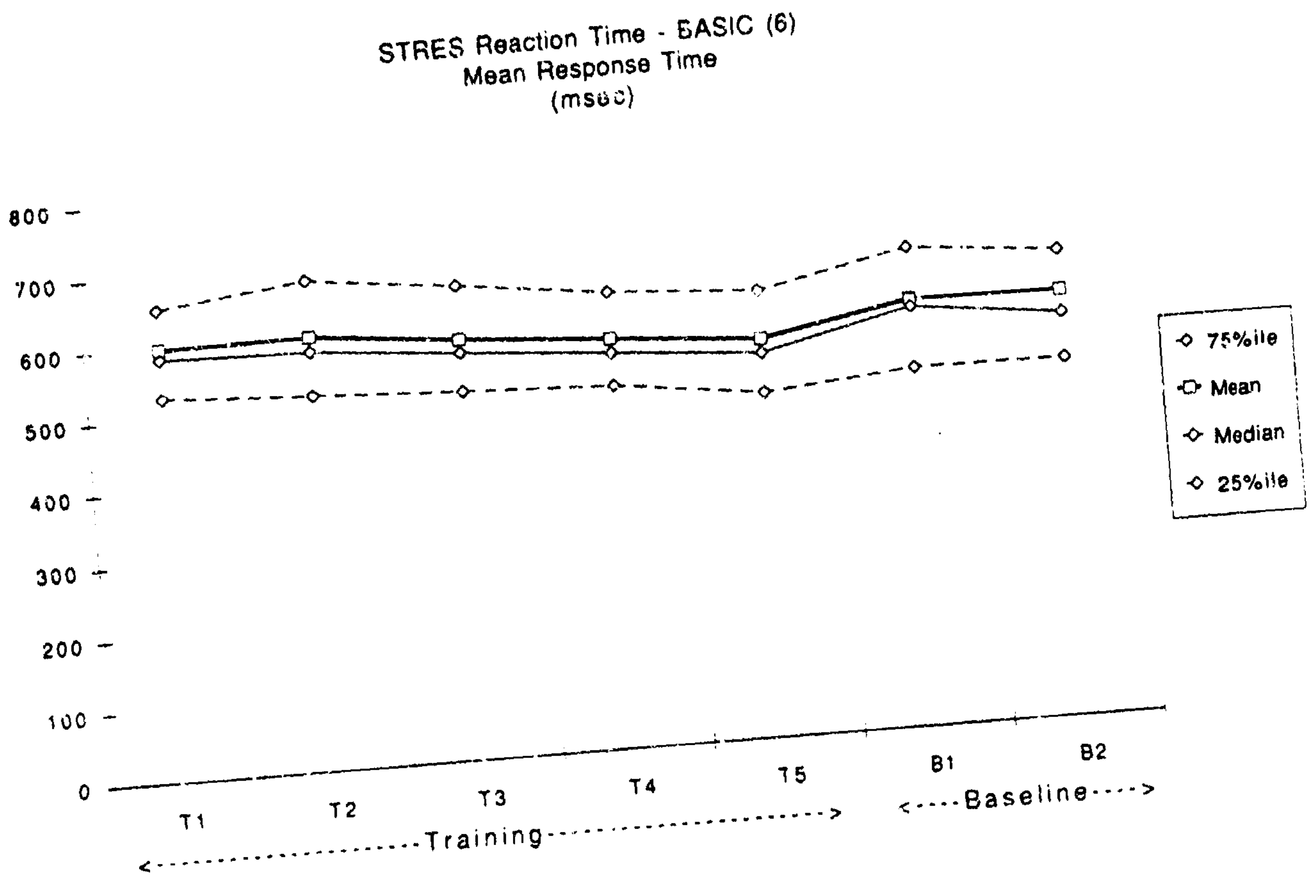
BLOCK=BASIC CODE=6

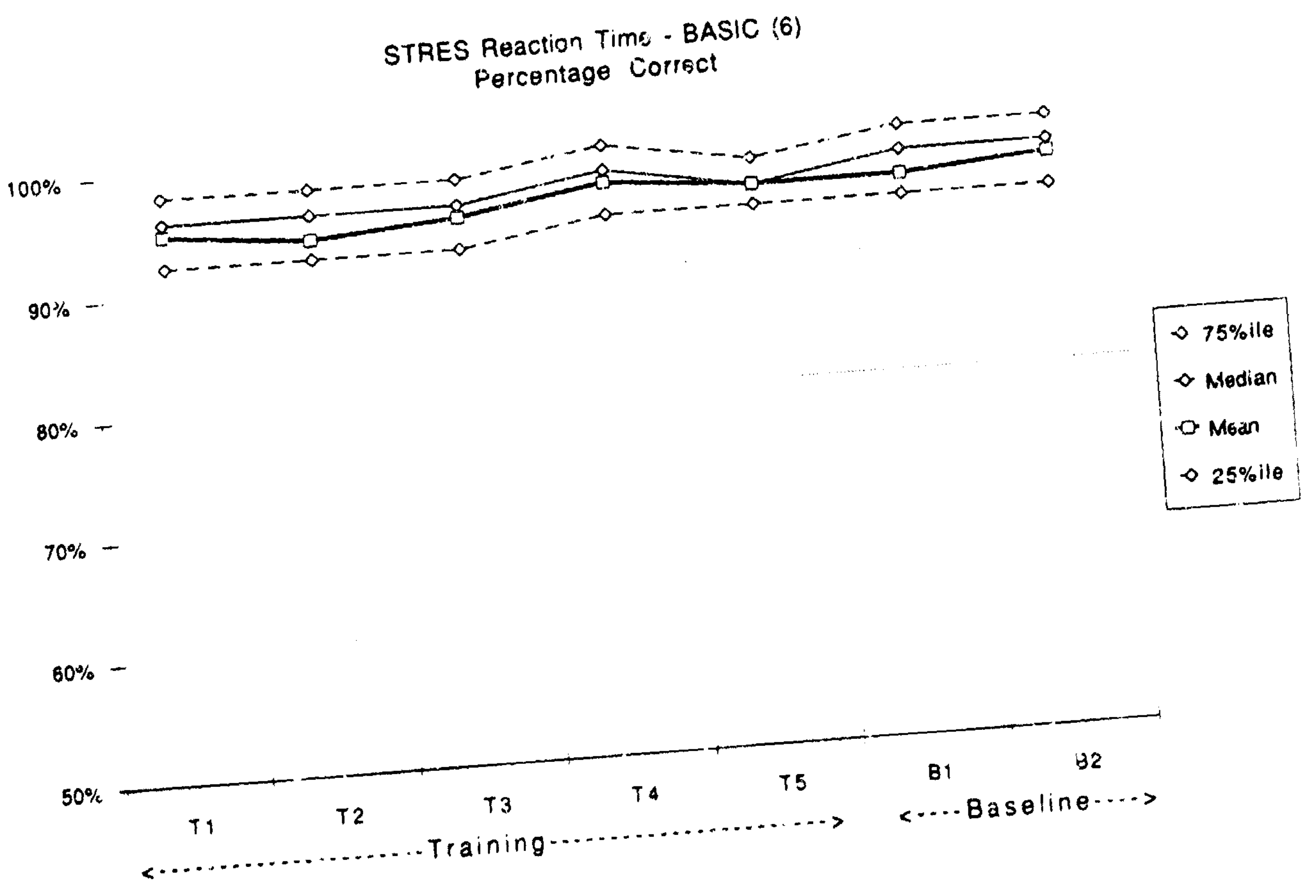
RTMN

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	599	504	695	96	585	530	655
T2	601	494	709	107	581	520	681
T3	582	487	677	95	564	510	658
T4	567	478	657	90	548	502	633
T5	551	463	640	88	532	477	619
B1	590	484	696	106	580	495	663
B2	588	464	711	124	557	494	645
21	613	488	739	126	581	509	705
22	588	467	709	121	558	512	632
26	583	480	686	103	568	497	653
27	609	481	736	127	608	487	659

RTPC

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	95%	90%	100%	5%	95%	92%	98%
T2	94%	89%	100%	6%	96%	92%	98%
T3	95%	92%	99%	4%	96%	92%	98%
T4	97%	93%	100%	4%	98%	94%	100%
T5	96%	92%	100%	4%	96%	94%	98%
B1	96%	93%	100%	3%	98%	94%	100%
B2	97%	93%	100%	4%	98%	94%	100%
21	97%	94%	100%	3%	98%	94%	98%
22	97%	94%	100%	3%	98%	96%	98%
26	97%	93%	100%	3%	98%	94%	100%
27	96%	92%	100%	4%	97%	94%	98%


STRES Reaction Time - BASIC (6)
Mean Response Time
(msec)
800
700
600
500
400
300
200
100
0
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------------>
<----Baseline---->
75%ile
Mean
Median
25%ile


STRES Reaction Time - BASIC (6)
Percentage Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------------>
<----Baseline---->
75%ile
Median
Mean
25%ile

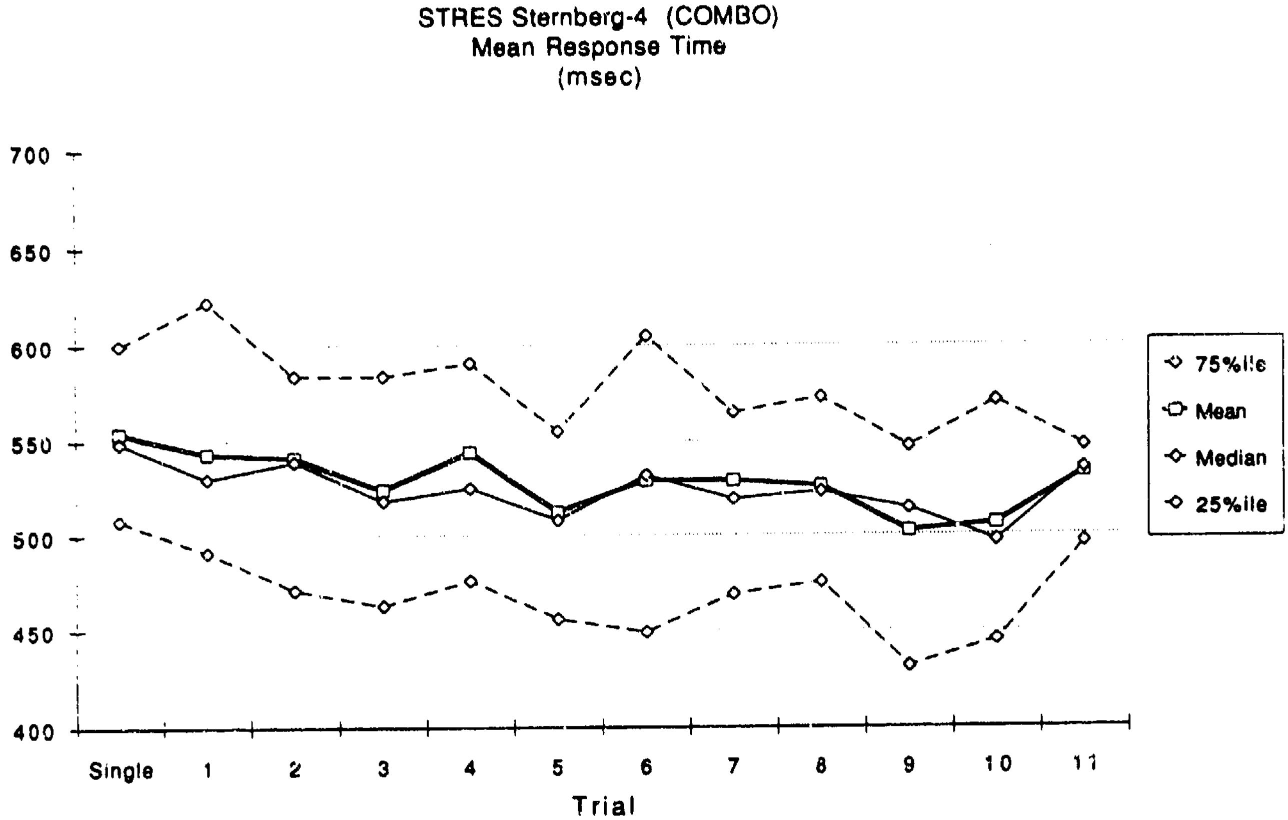
UNIVARIATE SUMMARY FOR STRES MEMORY SEARCH 4 - COMBO

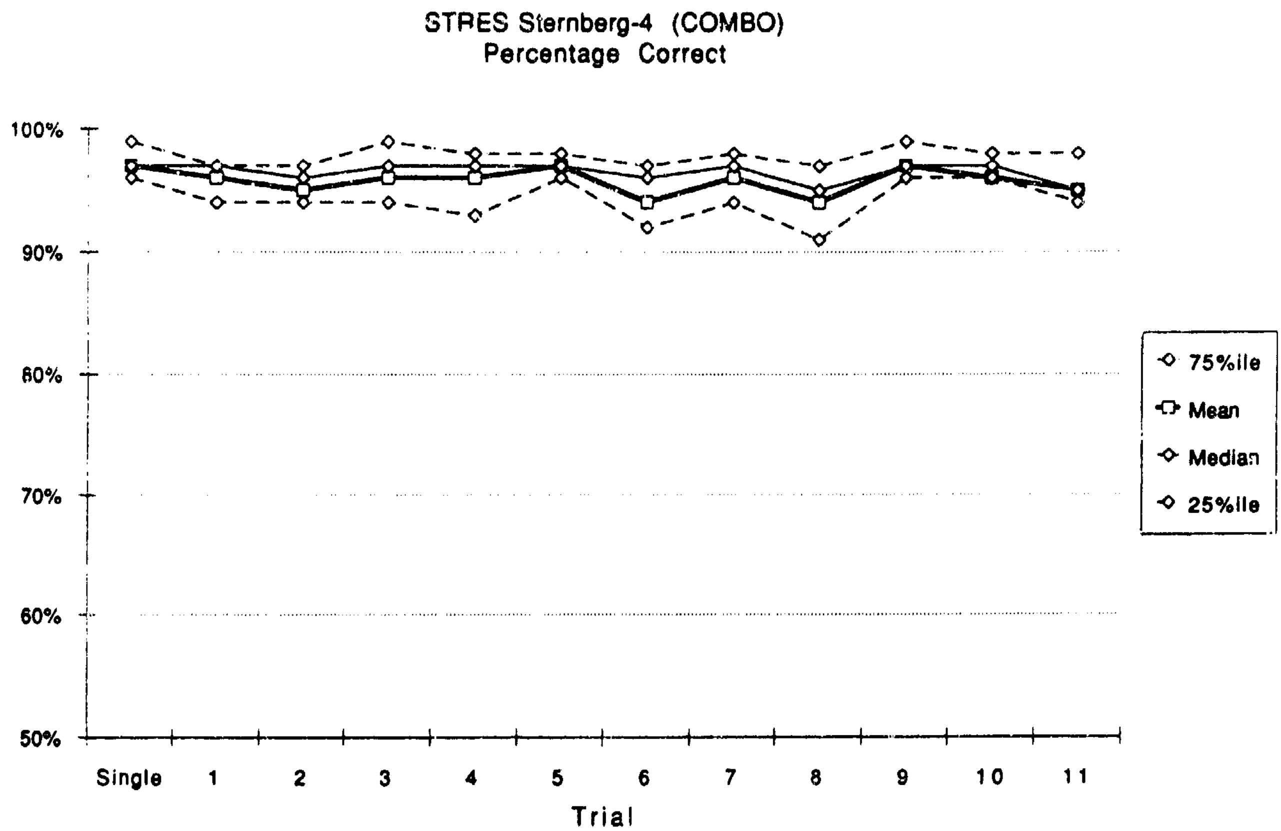
CBOMNO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
Single	554				549	508	600
1	543	460	625	83	530	491	622
2	541	459	623	82	539	471	584
3	524	434	614	90	518	463	584
4	544	456	633	88	525	476	591
5	512	423	602	89	508	456	555
6	529	437	620	92	532	449	605
7	529	453	605	76	519	469	565
8	526	449	603	77	523	475	573
9	502	431	572	70	514	431	547
10	506	432	581	74	497	445	571
11	533	462	604	71	535	496	547

CBOPCO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
Single	97%				97%	96%	99%
1	96%	93%	99%	3%	97%	94%	97%
2	95%	93%	98%	3%	96%	94%	97%
3	96%	93%	99%	3%	97%	94%	99%
4	96%	93%	99%	3%	97%	93%	98%
5	97%	94%	99%	3%	97%	96%	98%
6	94%	88%	99%	5%	96%	92%	97%
7	96%	92%	99%	4%	97%	94%	98%
8	94%	91%	98%	4%	95%	91%	97%
9	97%	95%	99%	2%	97%	96%	99%
10	96%	94%	99%	2%	97%	96%	98%
11	95%	92%	99%	3%	95%	94%	98%


STRES Sternberg-4 (COMBO)
Mean Response Time
(msec)
700
650
600
550
500
450
400
Single
1
2
3
4
5
6
7
8
9
10
11
Trial
75%ile
Mean
Median
25%ile


STRES Sternberg-4 (COMBO)
Percentage Correct
100%
90%
80%
70%
60%
50%
Single
1
2
3
4
5
6
7
8
9
10
11
Trial
75%ile
Mean
Median
25%ile

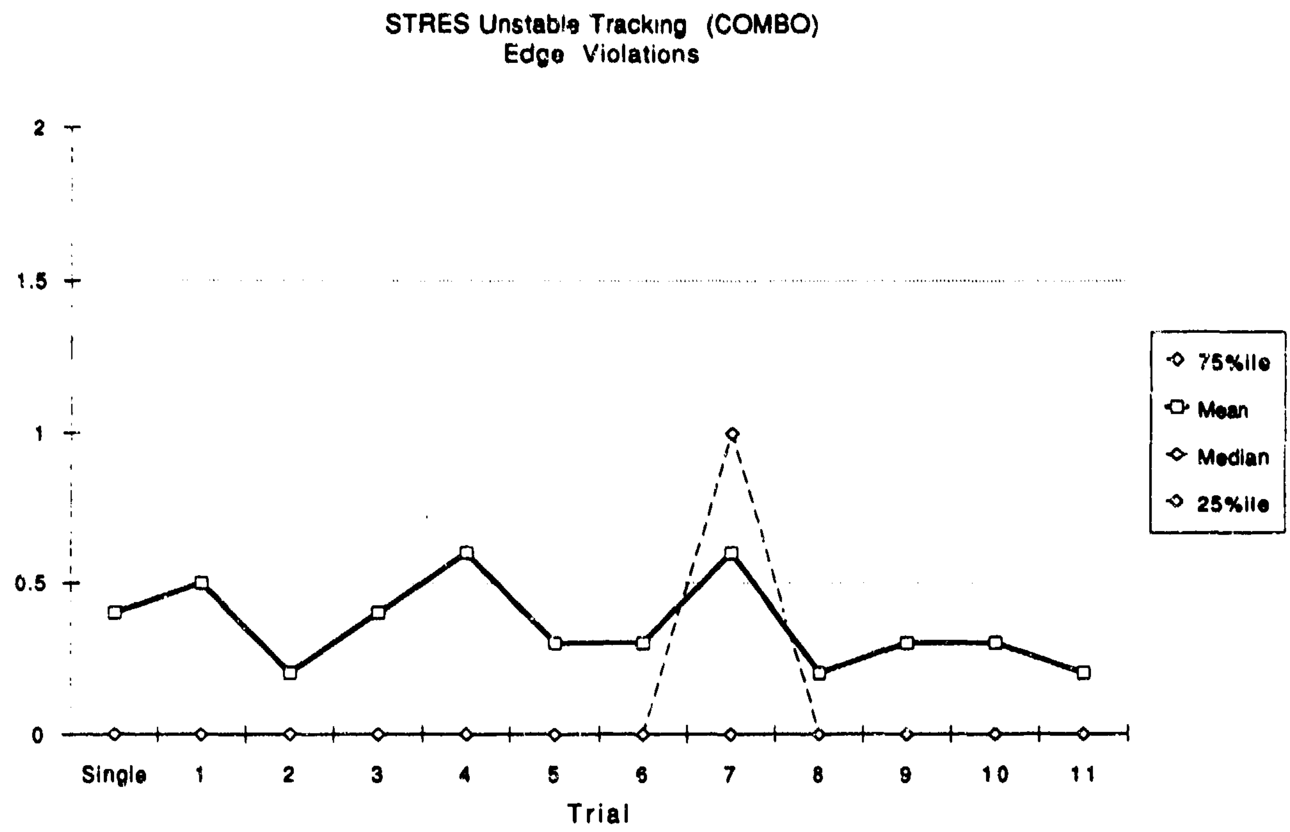
UNIVARIATE SUMMARY FOR STRES UNSTABLE TRACKING - COMBO

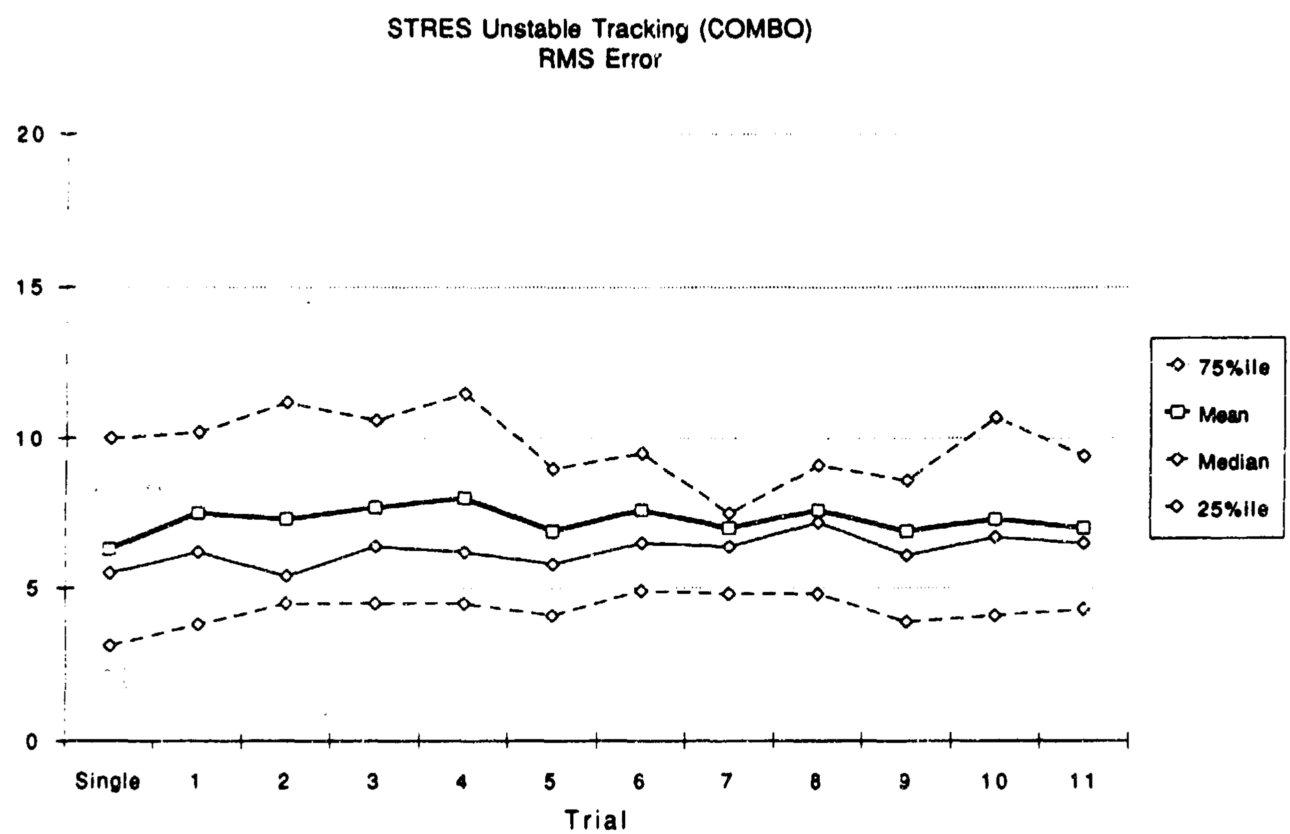
CBOEV

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
Single	0.4				0.0	0.0	0.0
1	0.5	0.0	1.8	1.2	0.0	0.0	0.0
2	0.2	0.0	0.9	0.7	0.0	0.0	0.0
3	0.4	0.0	1.3	0.9	0.0	0.0	0.0
4	0.6	0.0	2.8	2.1	0.0	0.0	0.0
5	0.3	0.0	1.1	0.9	0.0	0.0	0.0
6	0.3	0.0	1.3	1.0	0.0	0.0	0.0
7	0.6	0.0	1.6	1.1	0.0	0.0	1.0
8	0.2	0.0	0.7	0.5	0.0	0.0	0.0
9	0.3	0.0	1.7	1.4	0.0	0.0	0.0
10	0.3	0.0	1.2	0.9	0.0	0.0	0.0
11	0.2	0.0	1.0	0.8	0.0	0.0	0.0

CBORMS

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
Single	6.3				5.5	3.1	10.0
1	7.5	2.8	12.1	4.7	6.2	3.8	10.2
2	7.3	3.2	11.4	4.1	5.4	4.5	11.2
3	7.7	3.2	12.2	4.5	6.4	4.5	10.6
4	8.0	3.4	12.6	4.6	6.2	4.5	11.5
5	6.9	3.1	10.7	3.8	5.8	4.1	9.0
6	7.6	4.0	11.2	3.6	6.5	4.9	9.5
7	7.0	3.8	10.2	3.2	6.4	4.8	7.5
8	7.6	4.0	11.1	3.6	7.2	4.8	9.1
9	6.9	3.2	10.5	3.7	6.1	3.9	8.6
10	7.3	3.2	11.3	4.0	6.7	4.1	10.7
11	7.0	3.8	10.2	3.2	6.5	4.3	9.4


STRES Unstable Tracking (COMBO)
Edge Violations
2
1.5
1
0.5
0
Single
1
2
3
4
5
6
7
8
9
10
11
Trial
75%ile
Mean
Median
25%ile


STRES Unstable Tracking (COMBO)
RMS Error
20
15
10
5
0
Single
1
2
3
4
5
6
7
8
9
10
11
Trial
75%ile
Mean
Median
25%ile

APPENDIX D

CTS NORMATIVE DATA

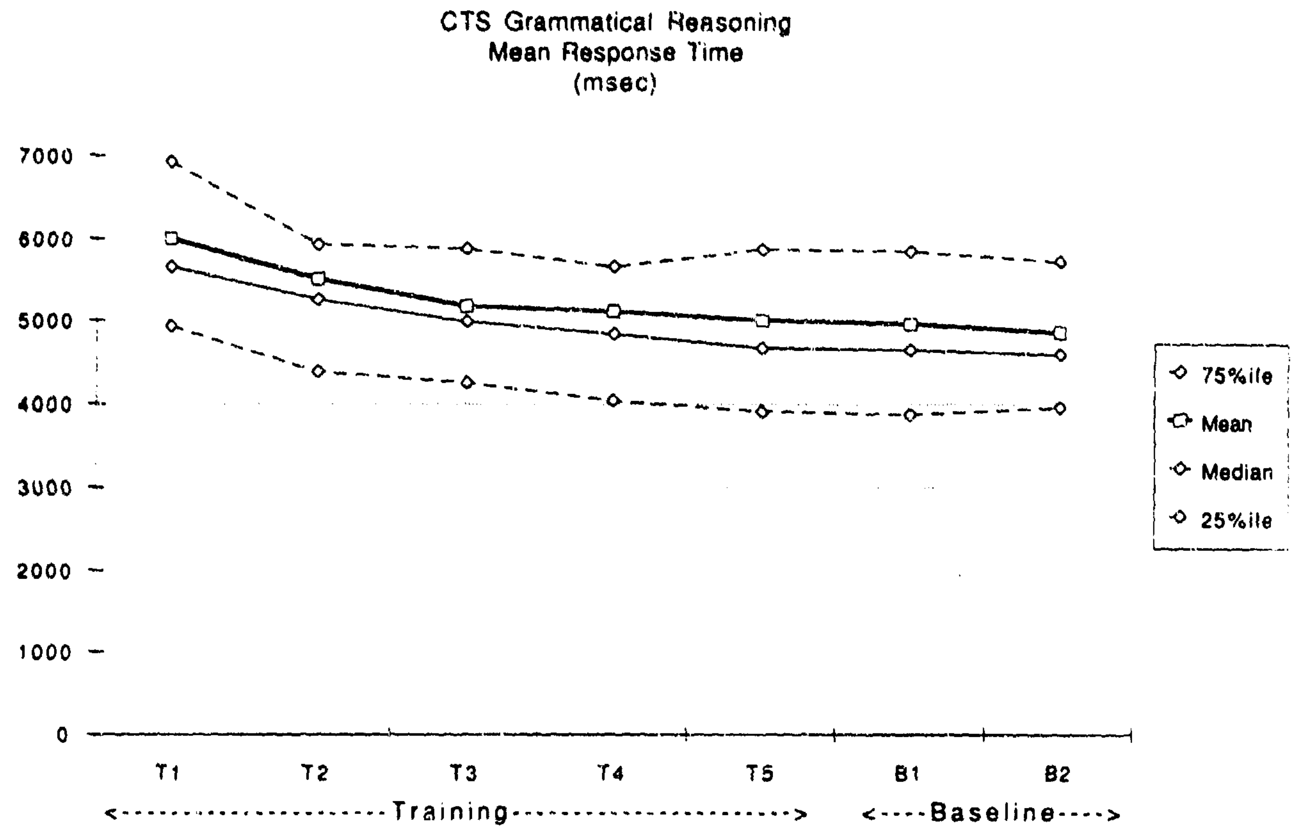
UNIVARIATE SUMMARY FOR CTS GRAMMATICAL REASONING

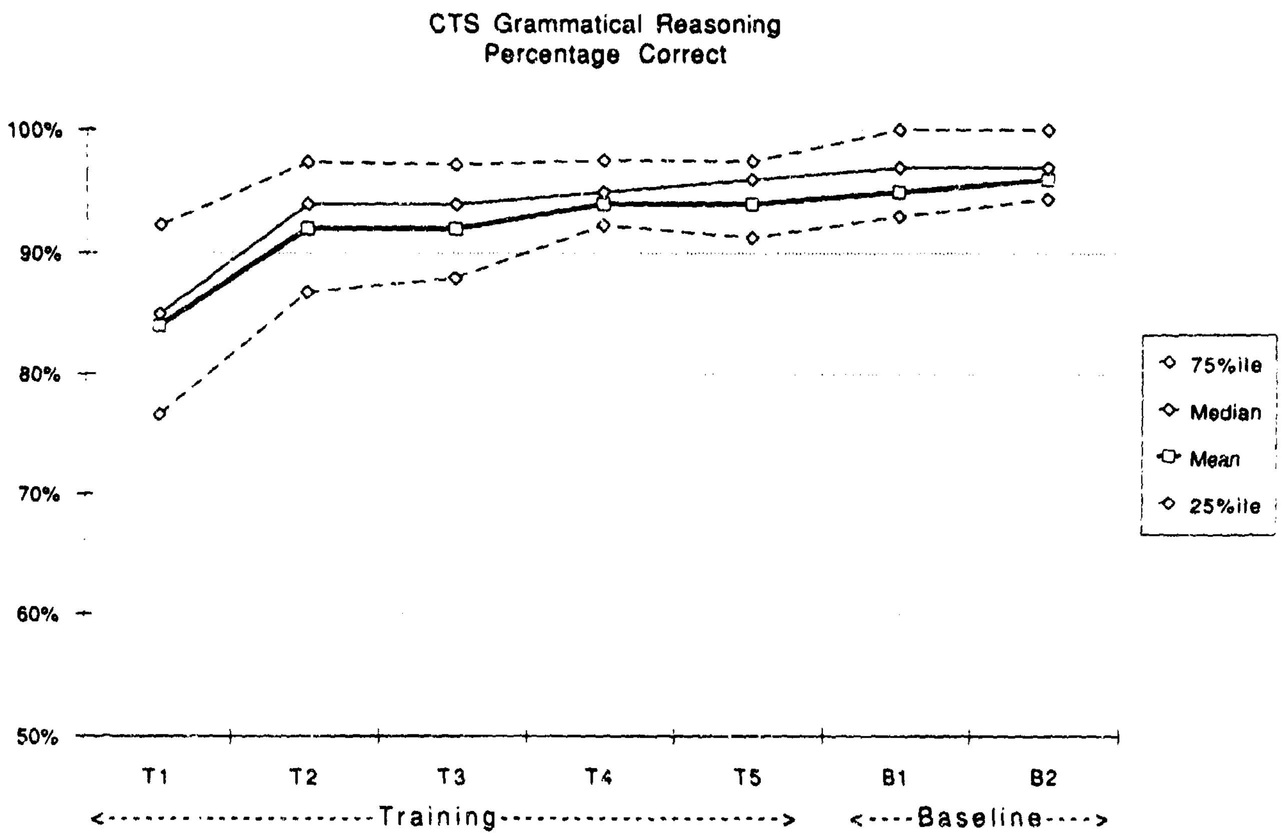
GRMNO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	5998	4527	7469	1471	5661	4934	6926
T2	5512	4032	6991	1480	5262	4391	5929
T3	5176	3716	6635	1460	4991	4261	5881
T4	5114	3616	6612	1498	4839	4047	5661
T5	5002	3523	6481	1479	4675	3914	5867
B1	4954	3528	6380	1426	4648	3876	5847
B2	4855	3401	6309	1454	4595	3958	5717
21	4948	3281	6616	1667	4644	3649	5797
22	4743	3348	6137	1395	4469	3833	5756
26	5311	3607	7015	1704	4969	4247	6132
27	4864	3651	6077	1213	4557	4197	5825

GRPOO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	84%	72%	96%	12%	85%	77%	92%
T2	92%	83%	100%	8%	94%	87%	97%
T3	92%	83%	100%	8%	94%	88%	97%
T4	94%	88%	100%	6%	95%	92%	98%
T5	94%	86%	100%	7%	96%	91%	98%
B1	95%	90%	100%	5%	97%	93%	100%
B2	96%	91%	100%	5%	97%	94%	100%
21	93%	86%	100%	7%	95%	89%	99%
22	96%	92%	100%	4%	97%	95%	100%
26	96%	89%	100%	7%	99%	95%	100%
27	96%	92%	100%	4%	97%	94%	100%


CTS Grammatical Reasoning
Mean Response Time
(msec)
7000
6000
5000
4000
3000
2000
1000
0
T1
T2
T3
T4
T5
B1
B2
<---------------------Training---------------------->
<----Baseline---->
75%ile
Mean
Median
25%ile


CTS Grammatical Reasoning
Percentage Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<---------------------Training---------------------->
<----Baseline---->
75%ile
Median
Mean
25%ile

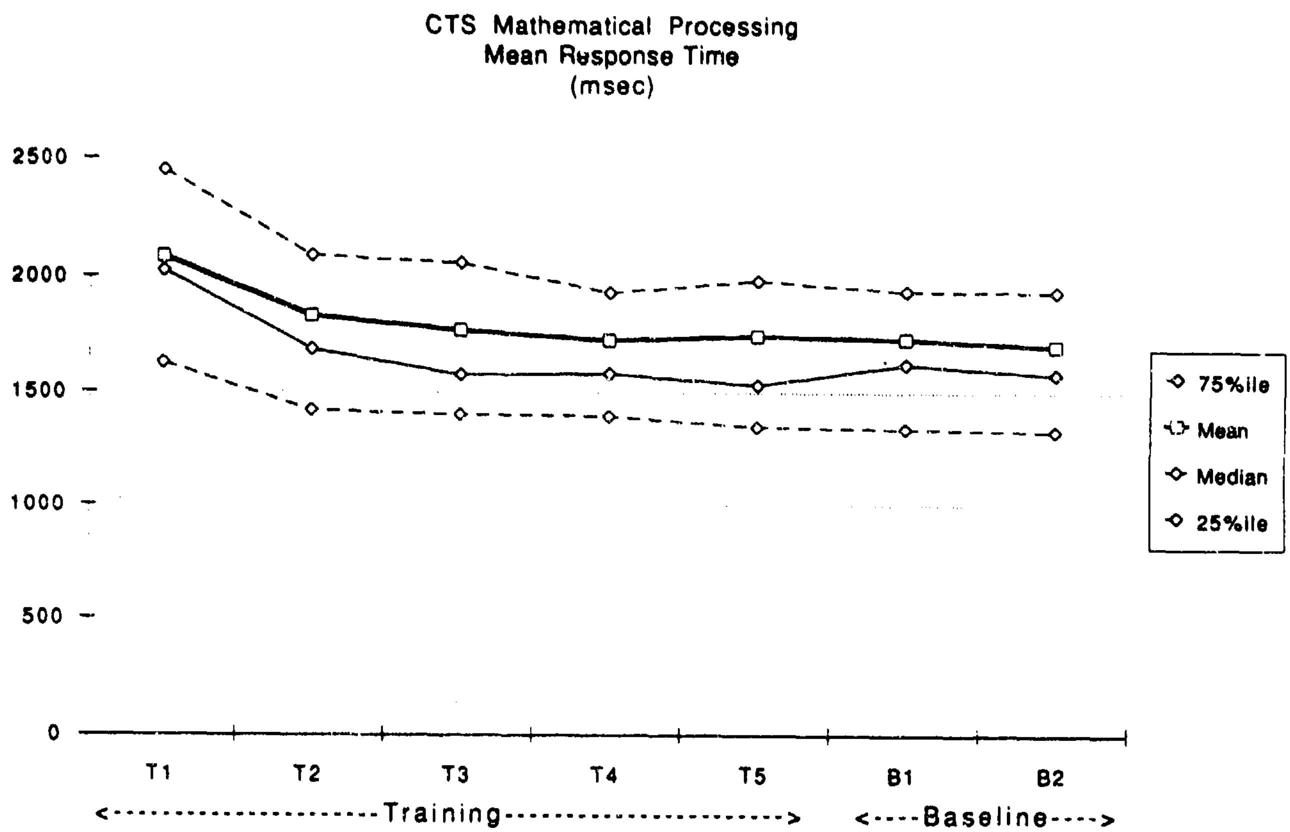
UNIVARIATE SUMMARY FOR CTS MATHEMATICAL PROCESSING

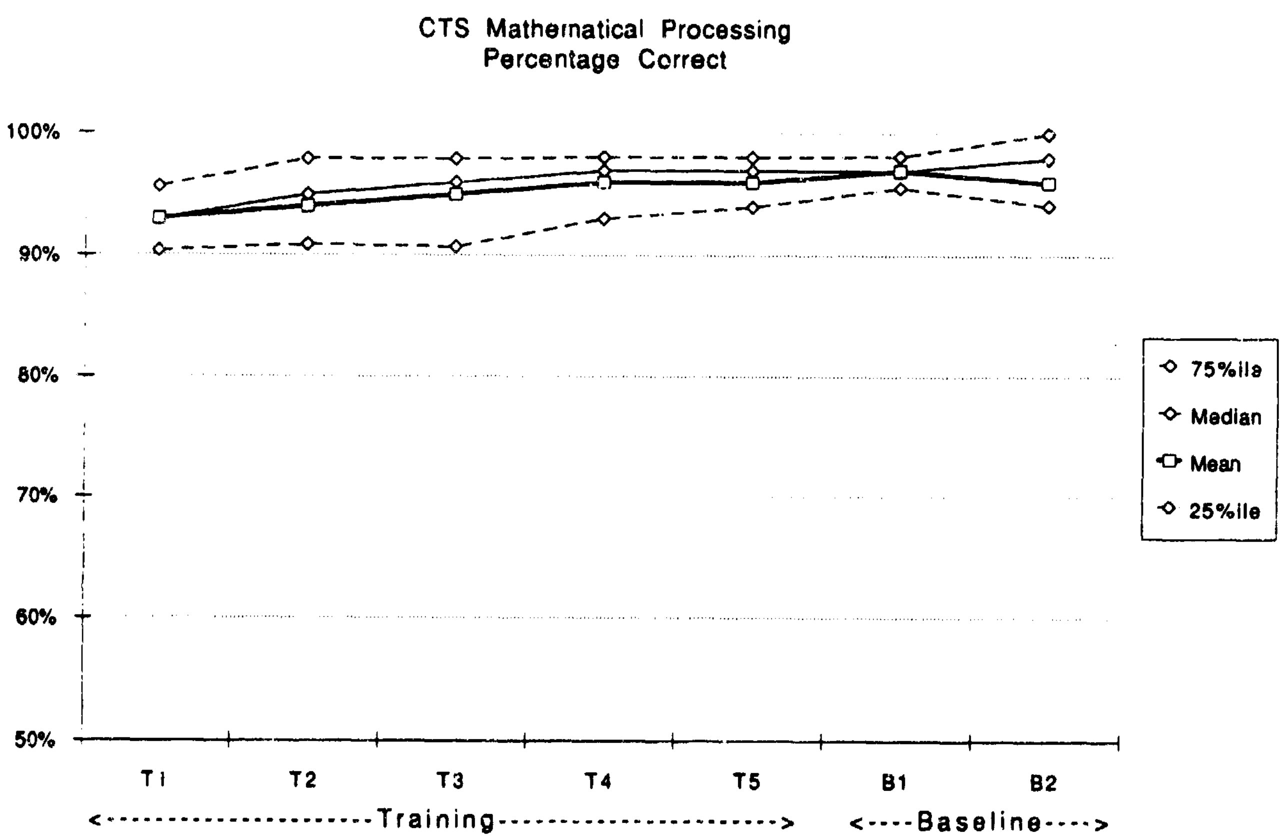
MPMNO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	2086	1534	2638	552	2026	1628	2449
T2	1830	1233	2427	597	1687	1422	2091
T3	1769	1236	2302	533	1578	1403	2060
T4	1726	1220	2233	507	1586	1398	1932
T5	1745	1161	2330	584	1535	1351	1984
B1	1732	1205	2260	527	1626	1341	1938
B2	1703	1201	2206	502	1581	1330	1935
21	1779	1259	2299	520	1685	1403	2032
22	1605	1215	1995	390	1549	1364	1809
26	1840	1335	2345	505	1650	1459	2266
27	1844	1295	2392	548	1695	1411	2320

MPPCO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	93%	88%	97%	5%	93%	90%	96%
T2	94%	89%	99%	5%	95%	91%	98%
T3	95%	90%	100%	5%	96%	91%	98%
T4	96%	92%	100%	4%	97%	93%	98%
T5	96%	93%	100%	3%	97%	94%	98%
B1	97%	94%	100%	3%	97%	96%	98%
B2	96%	92%	100%	5%	98%	94%	100%
21	98%	95%	100%	2%	98%	96%	100%
22	97%	94%	100%	3%	98%	96%	100%
26	95%	91%	100%	4%	96%	93%	98%
27	96%	93%	100%	3%	97%	95%	98%


CTS Mathematical Processing
Mean Response Time
(msec)
2500
2000
1500
1000
500
0
T1
T2
T3
T4
T5
B1
B2
<---------------------Training---------------------->
<----Baseline---->
75%ile
Mean
Median
25%ile


CTS Mathematical Processing
Percentage Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<---------------------Training---------------------->
<----Baseline---->
75%ile
Median
Mean
25%ile

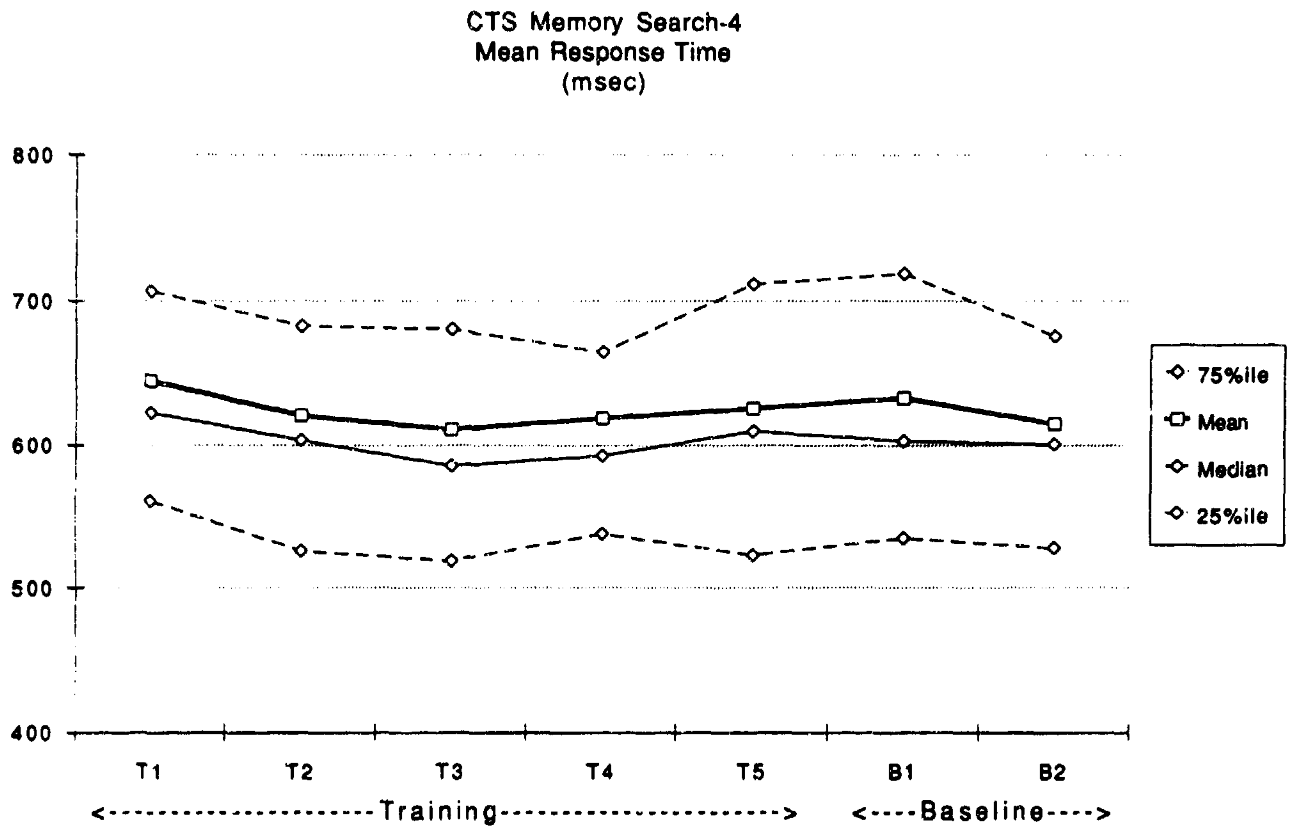
UNIVARIATE SUMMARY FOR CTS MEMORY SEARCH

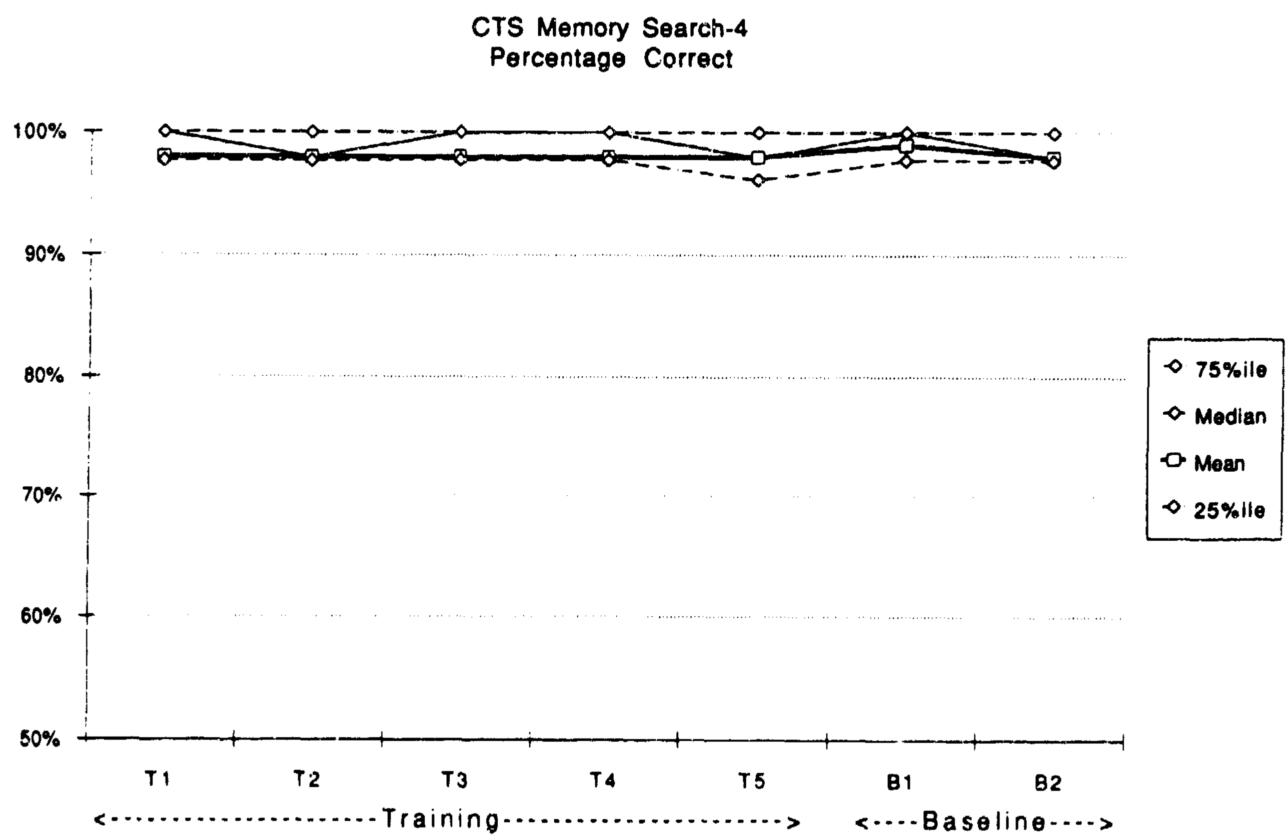
MSMNO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	645	540	749	105	623	561	707
T2	621	507	735	114	604	526	683
T3	611	497	725	114	586	519	681
T4	619	498	740	121	593	538	665
T5	626	499	754	127	610	523	712
B1	633	498	768	135	603	535	719
B2	615	498	733	118	601	528	676
21	663	532	794	131	658	586	705
22	642	509	774	133	593	547	691
26	675	543	807	132	641	576	788
27	664	526	801	137	622	558	796

MSPCO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	98%	95%	100%	3%	100%	98%	100%
T2	98%	96%	100%	3%	98%	98%	100%
T3	98%	95%	100%	3%	100%	98%	100%
T4	98%	96%	100%	2%	100%	98%	100%
T5	98%	95%	100%	3%	98%	96%	100%
B1	99%	96%	100%	3%	100%	98%	100%
B2	98%	96%	100%	2%	98%	98%	100%
21	98%	94%	100%	3%	100%	98%	100%
22	99%	98%	100%	1%	100%	98%	100%
26	98%	94%	100%	4%	98%	98%	100%
27	98%	95%	100%	3%	98%	98%	100%


CTS Memory Search-4
Mean Response Time
(msec)
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<--------------------Training--------------------->
<----Baseline---->
75%ile
Mean
Median
25%ile


CTS Memory Search-4
Percentage Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<--------------------Training--------------------->
<----Baseline---->
75%ile
Median
Mean
25%ile

UNIVARIATE SUMMARY FOR CTS SPATIAL PROCESSING

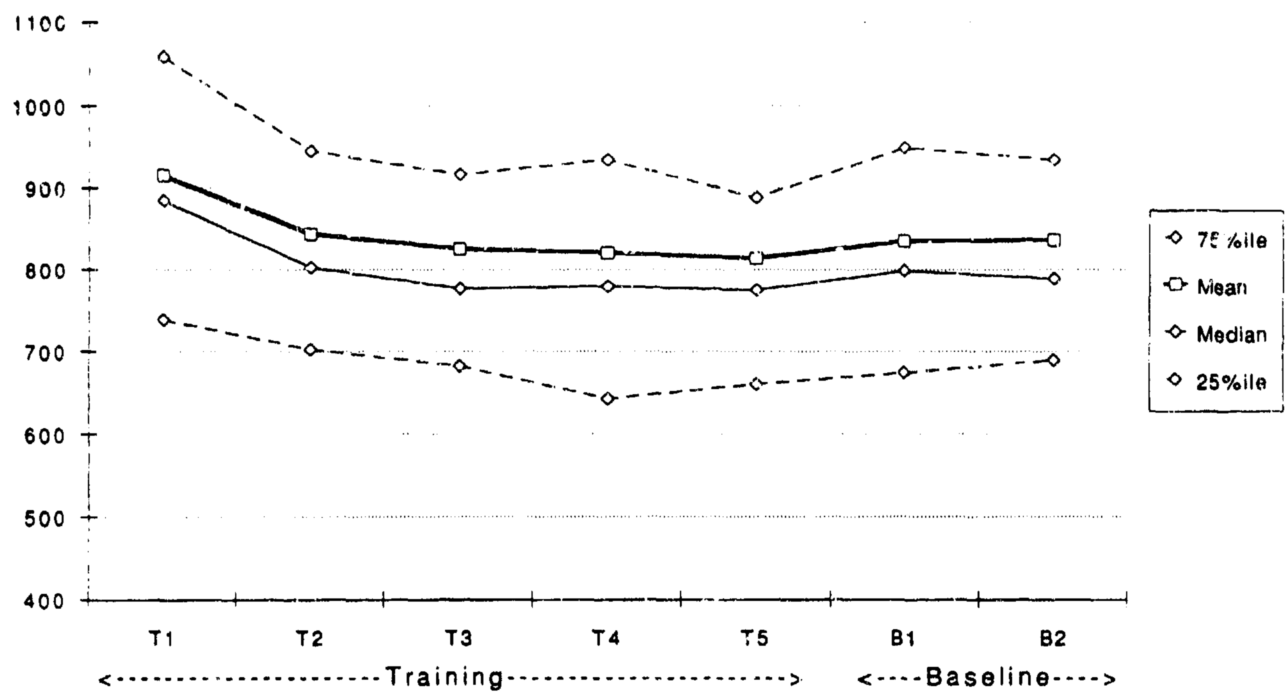
SPMNO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	916	681	1150	235	885	739	1059
T2	844	631	1057	213	803	703	946
T3	825	609	1041	216	777	683	917
T4	821	570	1072	251	780	643	935
T5	814	601	1027	213	775	661	889
B1	835	607	1062	227	799	675	949
B2	836	615	1056	220	789	690	934
21	830	653	1007	177	817	705	911
22	799	621	976	178	783	677	885
26	853	634	1073	219	815	714	960
27	854	625	1082	228	782	723	1026

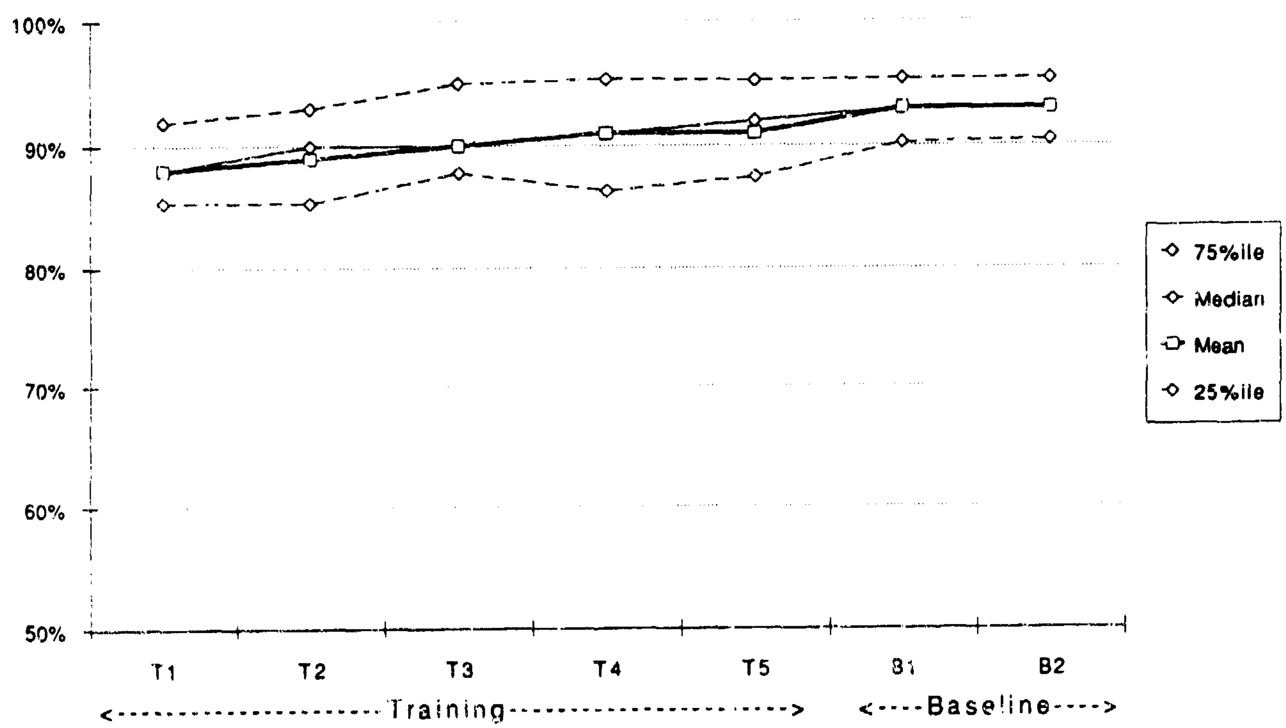
SPPCO

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	88%	82%	95%	6%	88%	85%	92%
T2	89%	83%	95%	6%	90%	85%	93%
T3	90%	84%	96%	6%	90%	88%	95%
T4	91%	85%	96%	6%	91%	86%	95%
T5	91%	86%	97%	5%	92%	88%	95%
B1	93%	89%	97%	4%	93%	90%	95%
B2	93%	89%	98%	4%	93%	90%	95%
21	93%	87%	98%	6%	95%	91%	95%
22	94%	89%	98%	5%	93%	92%	98%
26	93%	88%	98%	5%	93%	90%	95%
27	91%	85%	98%	6%	93%	88%	95%

CTS Spatial Processing
Mean Response Time
(msec)



CTS Spatial Processing
Percentage Correct



UNIVARIATE SUMMARY FOR CTS UNSTABLE TRACKING

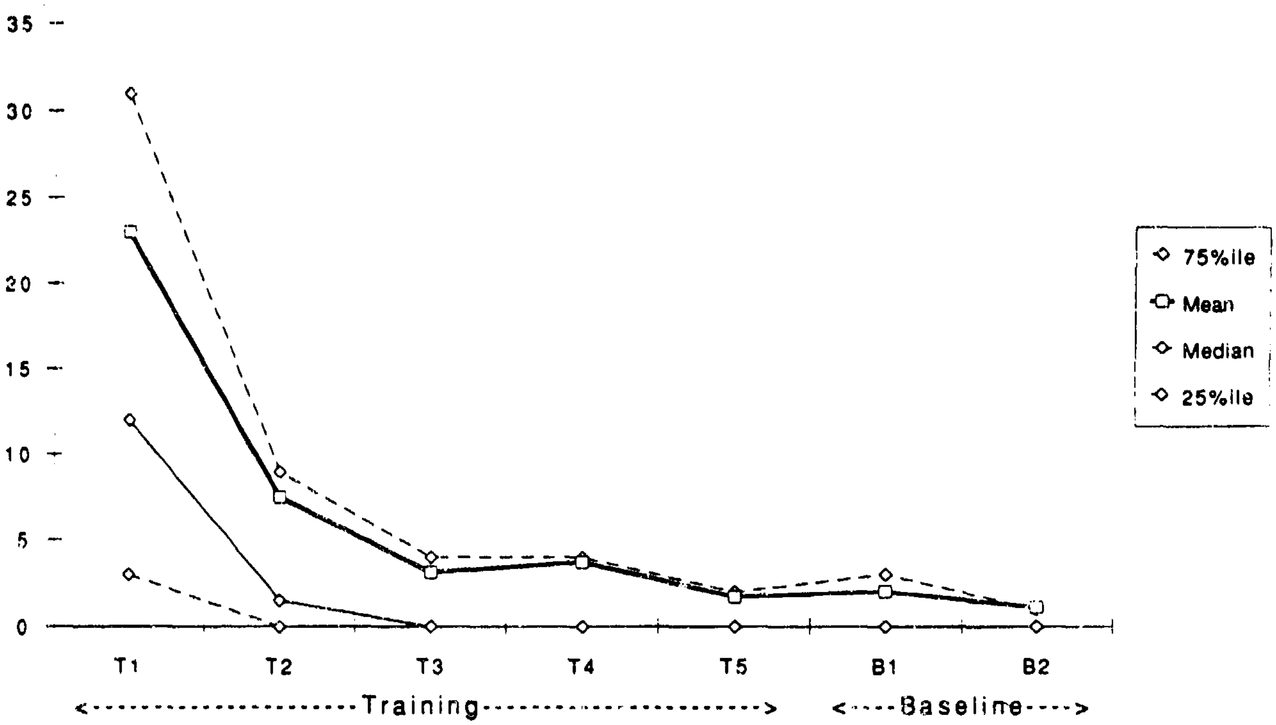
UTEV

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	23.0	0.0	[illegible]	28.2	12.0	3.0	31.0
T2	7.5	0.0	[illegible]	13.7	1.5	0.0	9.0
T3	3.1	0.0	[illegible]	5.7	0.0	0.0	4.0
T4	3.7	0.0	12.5	8.8	0.0	[illegible]	4.0
T5	1.7	0.0	5.3	3.6	0.0	0.0	2.0
B1	2.0	0.0	6.4	4.5	0.0	0.0	3.0
B2	1.1	0.0	3.8	2.6	0.0	0.0	1.0
21	1.9	0.0	5.7	3.7	0.0	0.0	2.0
22	2.1	0.0	5.7	3.6	0.0	0.0	4.0
26	1.2	0.0	3.6	2.3	0.0	0.0	2.0
27	1.3	0.0	4.3	3.1	0.0	0.0	0.0

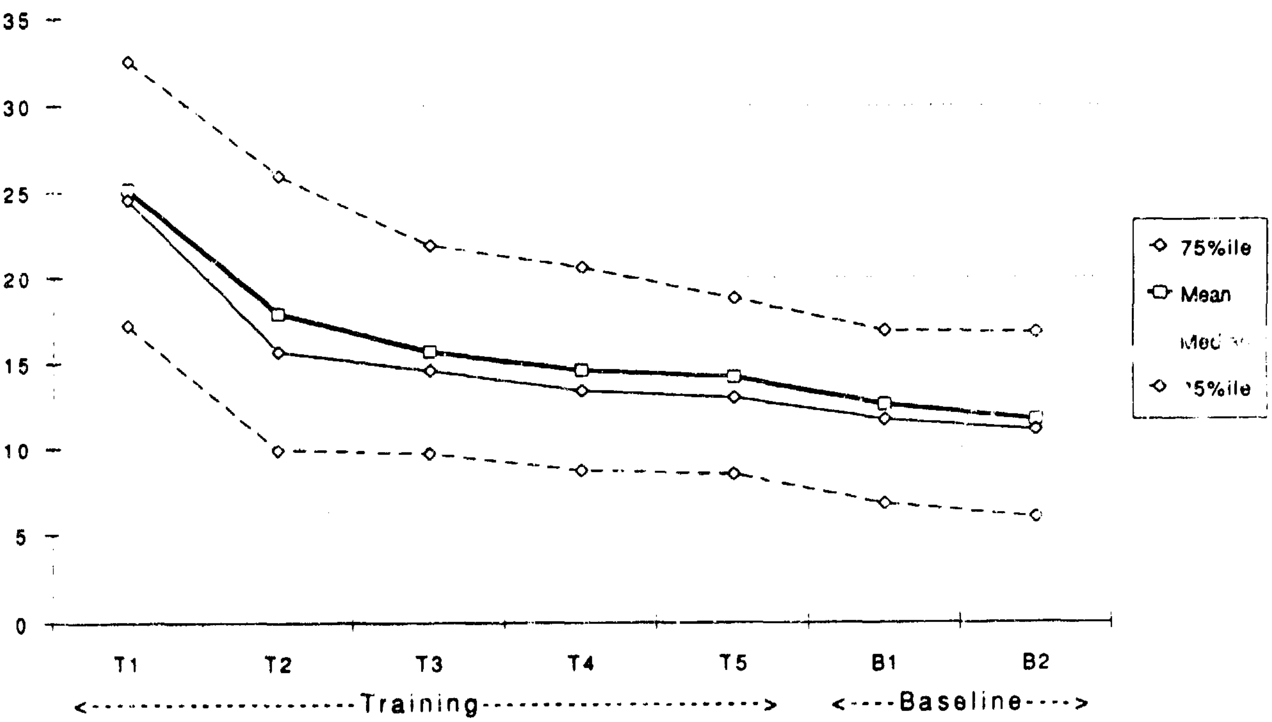
UTRMS

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	25.2	15.2	35.3	10.0	24.6	17.2	32.6
T2	17.9	8.2	27.7	9.8	15.7	9.9	26.0
T3	15.7	8.1	23.3	7.6	14.6	9.7	21.9
T4	14.6	6.6	22.5	7.9	13.4	8.7	20.6
T5	14.2	6.8	21.5	7.4	13.0	8.5	18.8
B1	12.6	5.2	20.0	7.4	11.7	6.8	16.9
B2	11.7	4.8	18.7	7.0	11.1	6.0	16.8
21	13.5	6.6	20.3	6.9	11.9	9.9	15.3
22	12.3	5.2	19.5	7.1	10.7	6.3	18.0
26	13.3	6.1	20.5	7.2	10.5	7.5	17.5
27	13.2	7.2	19.2	6.0	11.7	9.0	17.5

CTS Unstable Tracking
Edge Violations



CTS Unstable Tracking
RMS Error



APPENDIX E

WRPAB NORMATIVE DATA

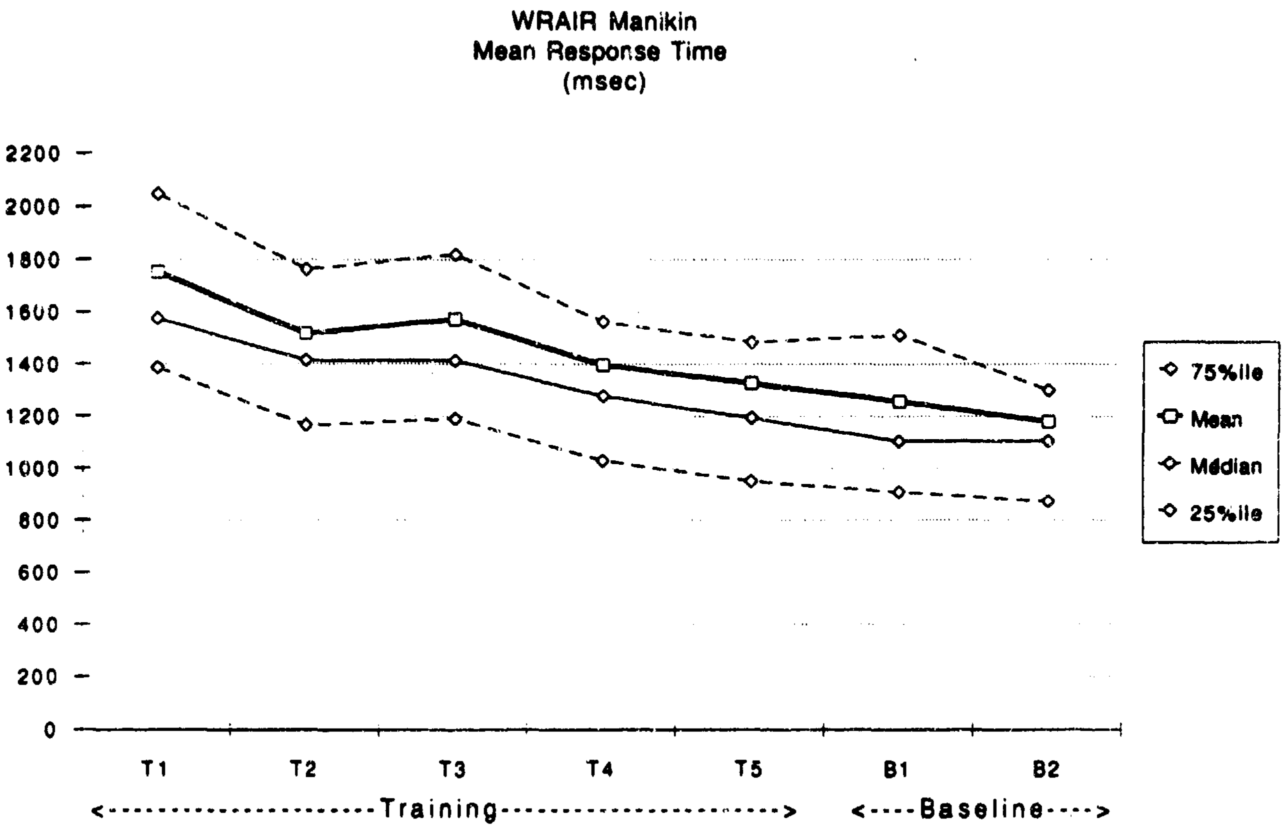
UNIVARIATE SUMMARY FOR WRAIR MANIKIN

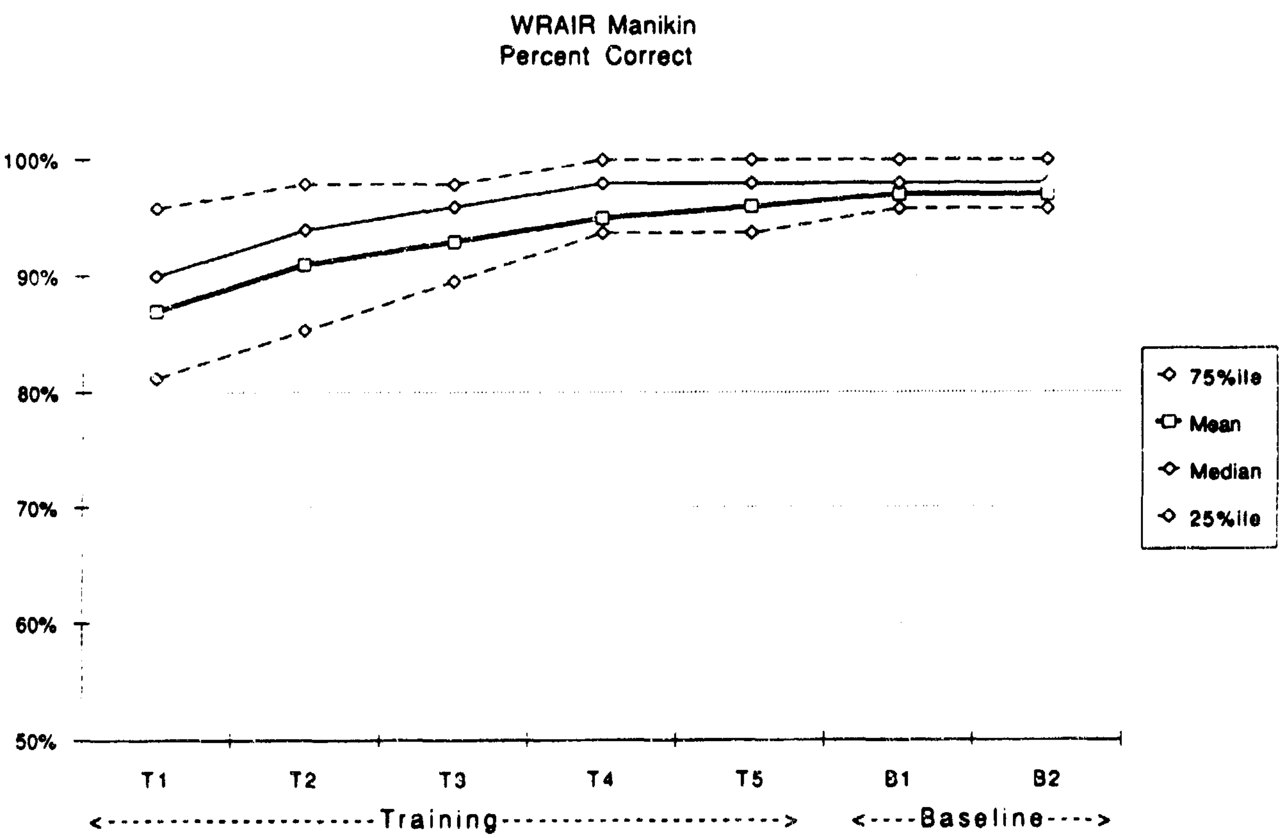
MANMN

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	1754	1159	2349	595	1575	1387	2049
T2	1517	950	2083	567	1416	1167	1765
T3	1573	907	2238	665	1415	1191	1821
T4	1396	811	1980	585	1279	1031	1563
T5	1328	748	1908	580	1195	951	1485
B1	1257	735	1780	522	1104	909	1513
B2	1179	713	1646	466	1105	872	1301
21	1272	805	1738	466	1165	928	1529
22	1171	744	1598	427	1062	858	1393
26	1220	625	1815	595	1083	830	1268
27	1048	512	1584	536	966	728	1147

MANPC

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	87%	76%	99%	11%	90%	81%	96%
T2	91%	82%	100%	9%	94%	85%	98%
T3	93%	85%	100%	8%	96%	90%	98%
T4	95%	90%	100%	6%	98%	94%	100%
T5	96%	91%	100%	5%	98%	94%	100%
B1	97%	94%	100%	3%	98%	96%	100%
B2	97%	93%	100%	4%	98%	96%	100%
21	97%	94%	100%	3%	98%	96%	100%
22	96%	92%	100%	4%	98%	94%	100%
26	98%	95%	100%	3%	98%	96%	100%
27	97%	94%	100%	4%	98%	96%	100%


WRAIR Manikin
Mean Response Time
(msec)
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
T1
T2
T3
T4
T5
B1
B2
<---------------------Training--------------------->
<----Baseline---->
75%ile
Mean
Median
25%ile


WRAIR Manikin
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<---------------------Training--------------------->
<----Baseline---->
75%ile
Mean
Median
25%ile

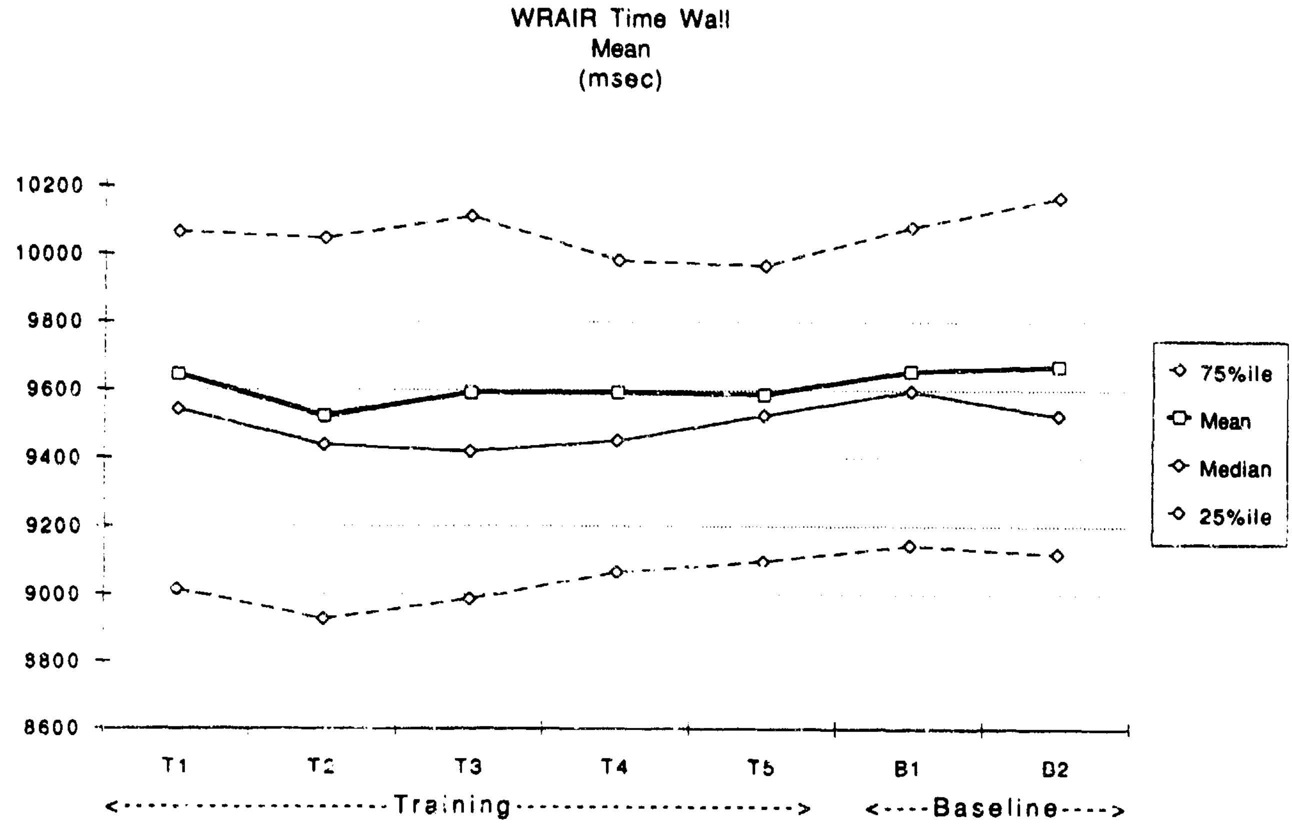
UNIVARIATE SUMMARY FOR WRAIR TIME WALL

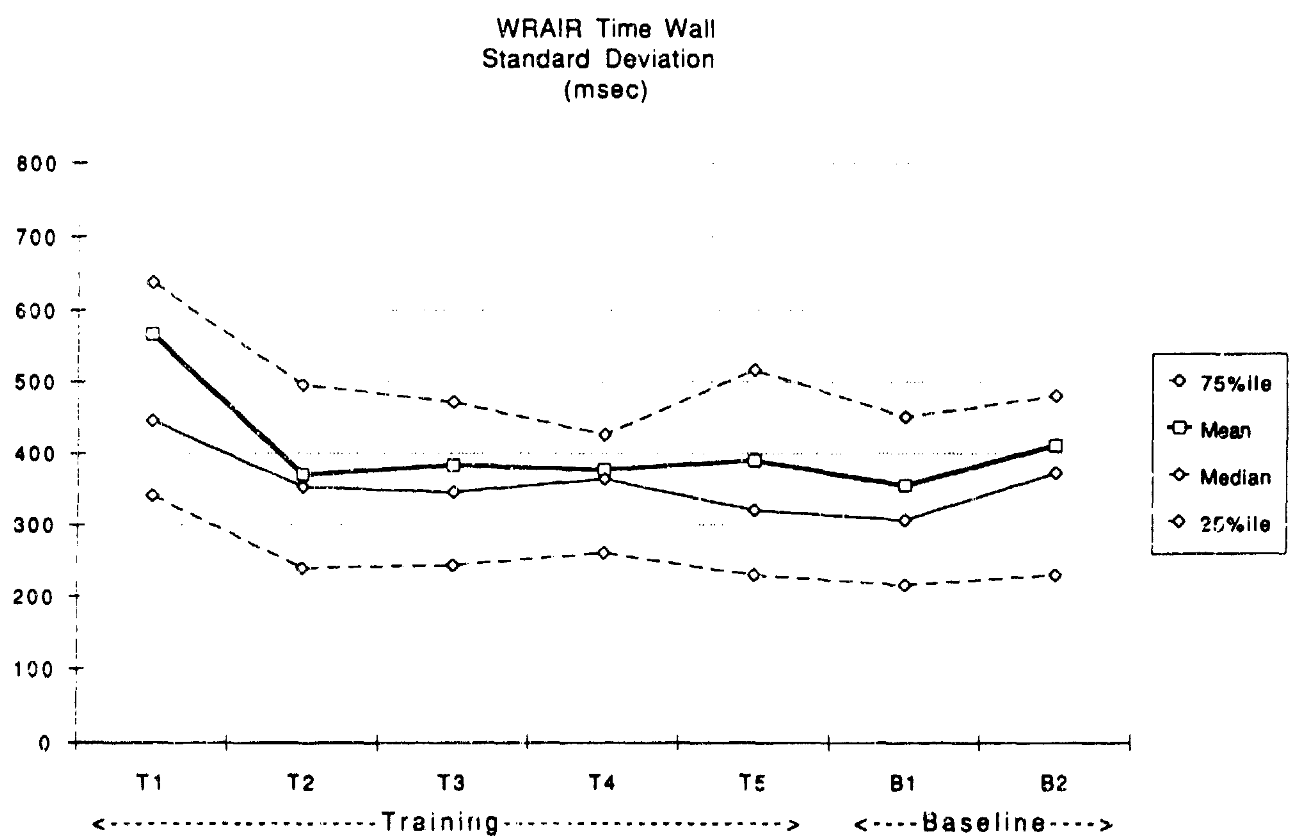
TIMMN

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	9646	8808	10485	839	9543	9016	10066
T2	9524	8746	10303	778	9440	8927	10048
T3	9591	8809	10373	782	9420	8986	10112
T4	9592	8745	10439	847	9453	9066	9981
T5	9587	8830	10343	757	9526	9098	9967
B1	9655	8852	10457	803	9595	9146	10079
B2	9670	8802	10538	868	9525	9121	10167
21	9818	8704	10931	1114	9597	9107	10263
22	9823	8724	10923	1099	9486	9181	10191
26	9486	8652	10320	834	9575	8691	9925
27	9636	8748	10524	888	9639	9058	9915

TIMSD

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	567	140	994	427	446	341	639
T2	370	202	537	168	353	240	496
T3	383	173	593	210	346	244	472
T4	377	203	551	174	364	261	426
T5	390	160	621	230	320	230	517
B1	355	164	546	191	306	216	451
B2	411	155	668	256	373	230	481
21	426	231	621	195	371	294	516
22	390	151	629	239	327	215	439
26	390	168	612	222	322	262	430
27	425	229	620	195	411	256	572


WRAIR Time Wall
Mean
(msec)
10200
10000
9800
9600
9400
9200
9000
8800
8600
75%ile
Mean
Median
25%ile
T1
T2
T3
T4
T5
B1
B2
<----------Training---------->
<----Baseline---->


WRAIR Time Wall
Standard Deviation
(msec)
800
700
600
500
400
300
200
100
0
75%ile
Mean
Median
25%ile
T1
T2
T3
T4
T5
B1
B2
<----------Training---------->
<----Baseline---->

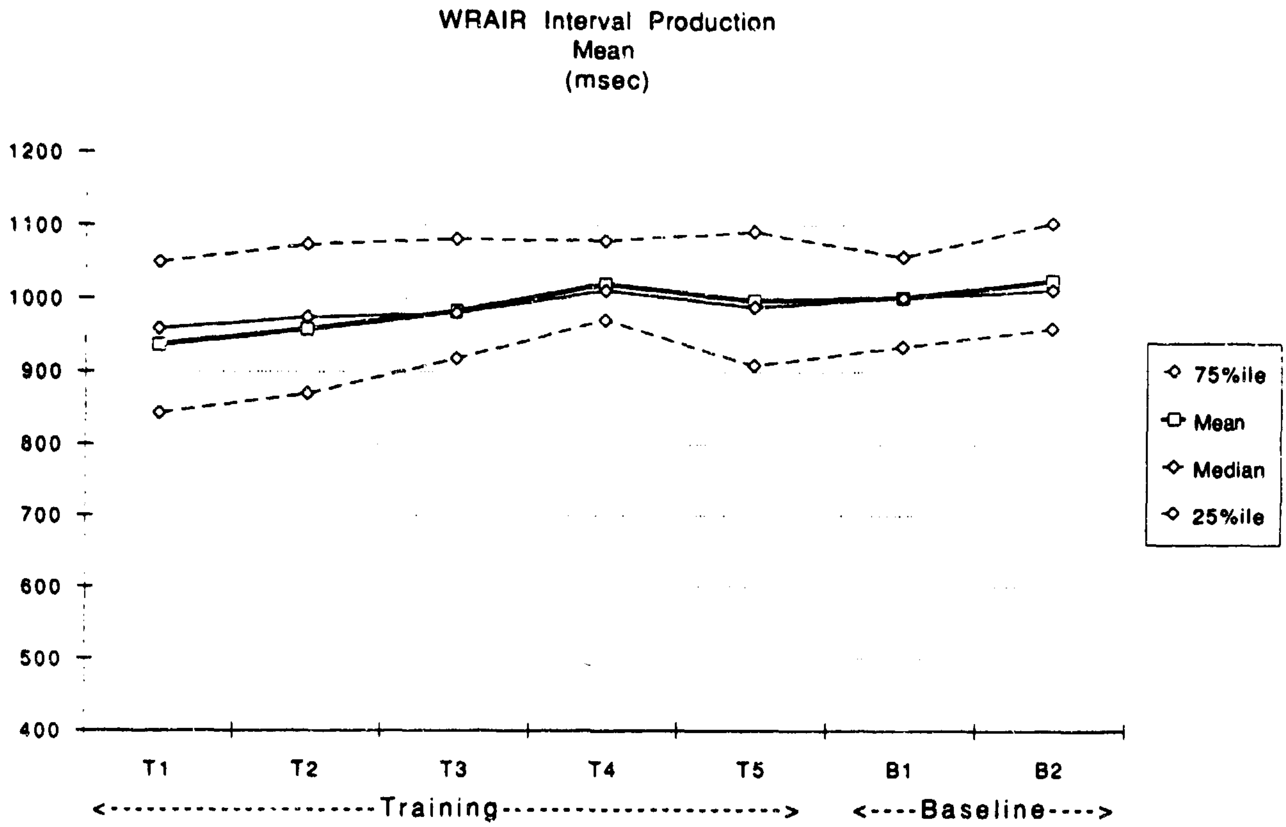
UNIVARIATE SUMMARY FOR WRAIR INTER /AL PRODUCTION

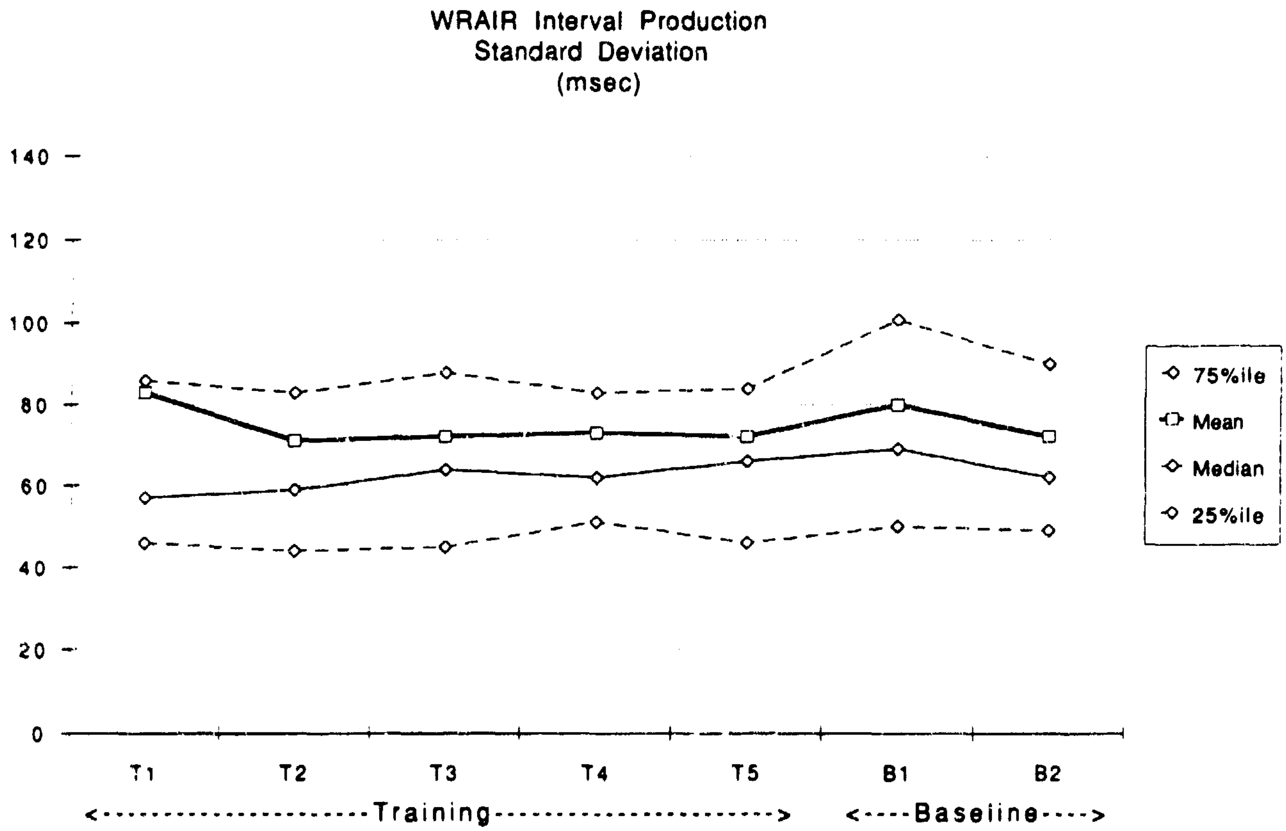
INTMN

Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	937	745	1128	192	959	844	1050
T2	958	814	1102	144	974	870	1074
T3	982	841	1122	141	980	918	1081
T4	1020	919	1120	101	1011	971	1079
T5	997	861	1133	136	988	909	1091
B1	1001	872	1130	129	1001	935	1057
B2	1024	895	1152	129	1011	959	1102
21	1009	883	1135	126	985	923	1050
22	1019	908	1130	111	1014	972	1081
26	981	869	1093	112	983	919	1044
27	1011	843	1179	168	1030	957	1097

!NTSD

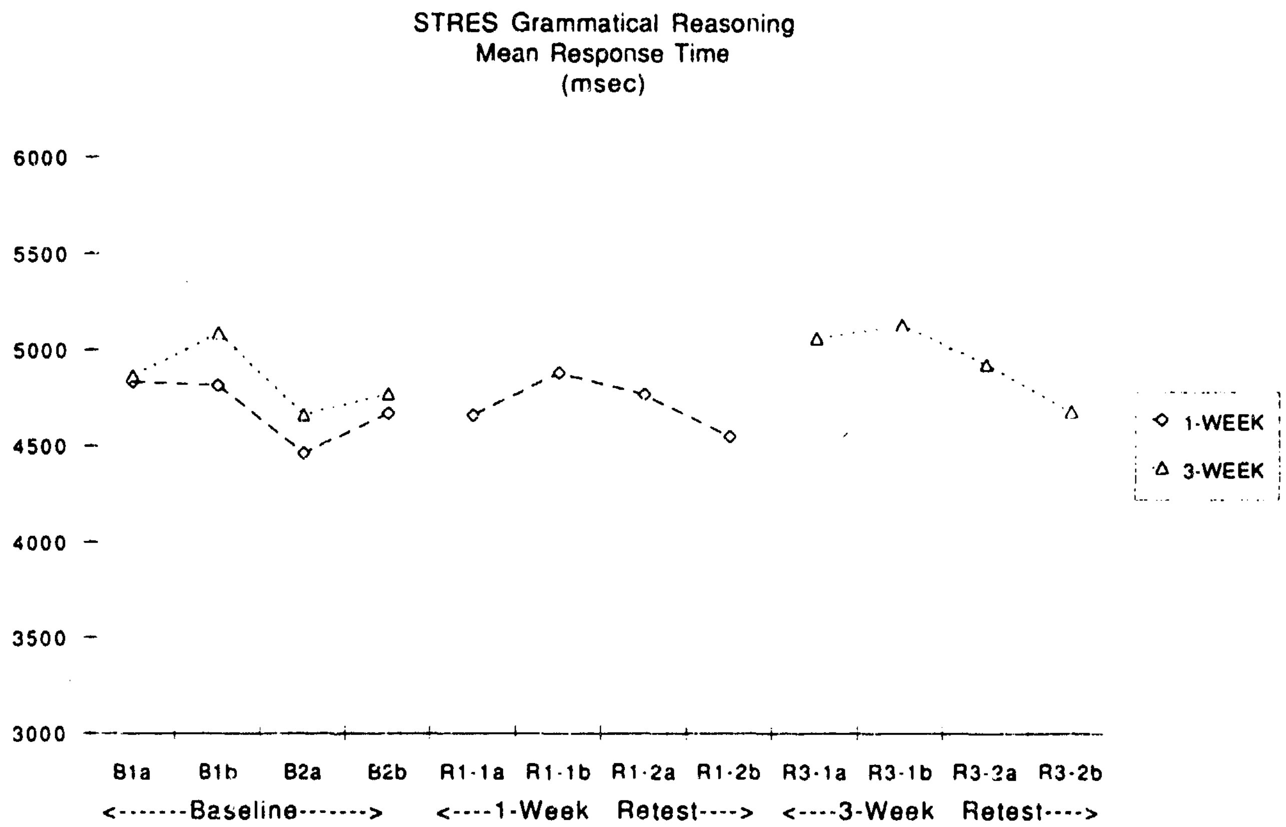
Session	Mean	Mean - 1 s.d.	Mean + 1 s.d.	Std. Dev.	Median	25%ile	75%ile
T1	83	0	166	83	57	46	86
T2	71	33	109	38	59	44	83
T3	72	36	109	36	64	45	88
T4	73	39	108	34	62	51	83
T5	72	37	107	35	66	46	84
B1	80	41	119	39	69	50	101
B2	72	40	105	33	62	49	90
21	89	53	124	36	81	68	97
22	91	46	135	45	75	60	110
26	76	32	119	44	68	55	82
27	73	34	111	38	66	55	82

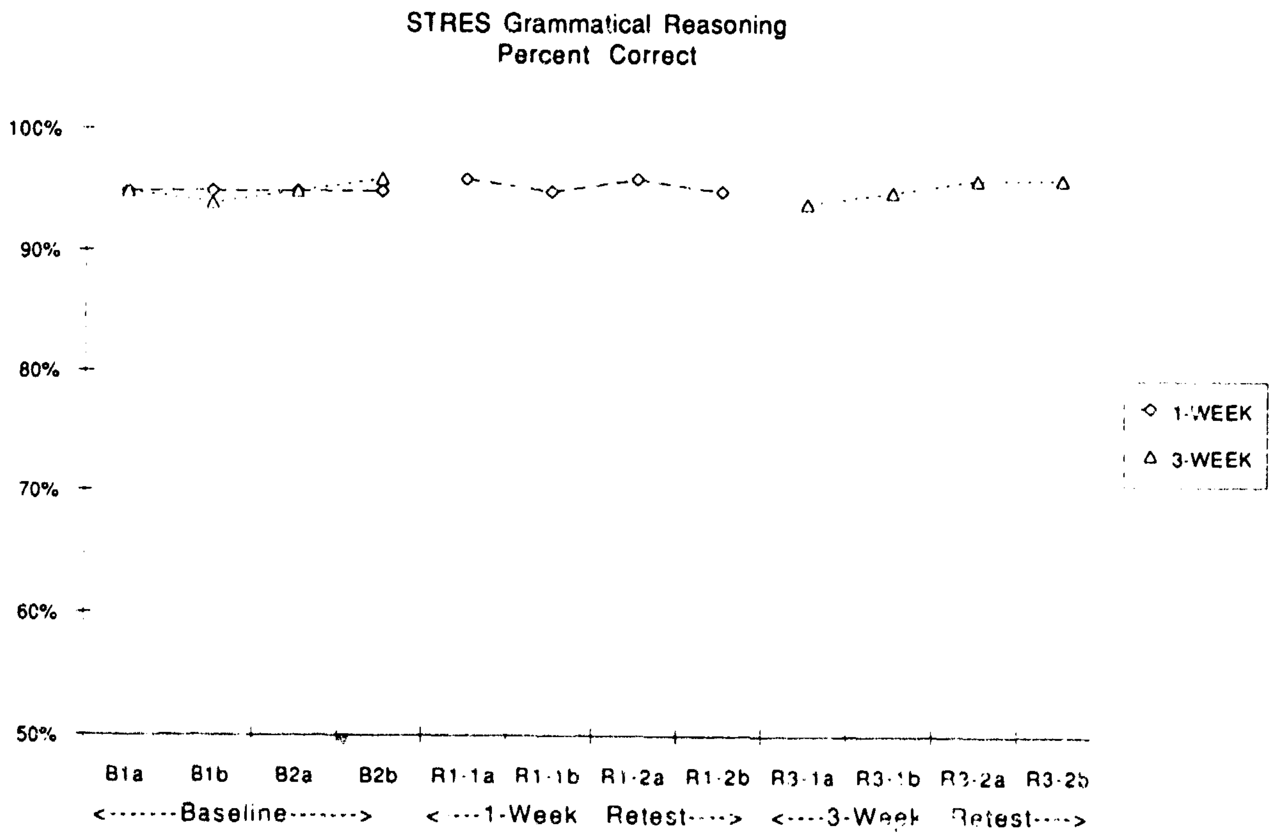

WRAIR Interval Production
Mean
(msec)
1200
1100
1000
900
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<---------------------Training---------------------->
<----Baseline---->
75%ile
Mean
Median
25%ile

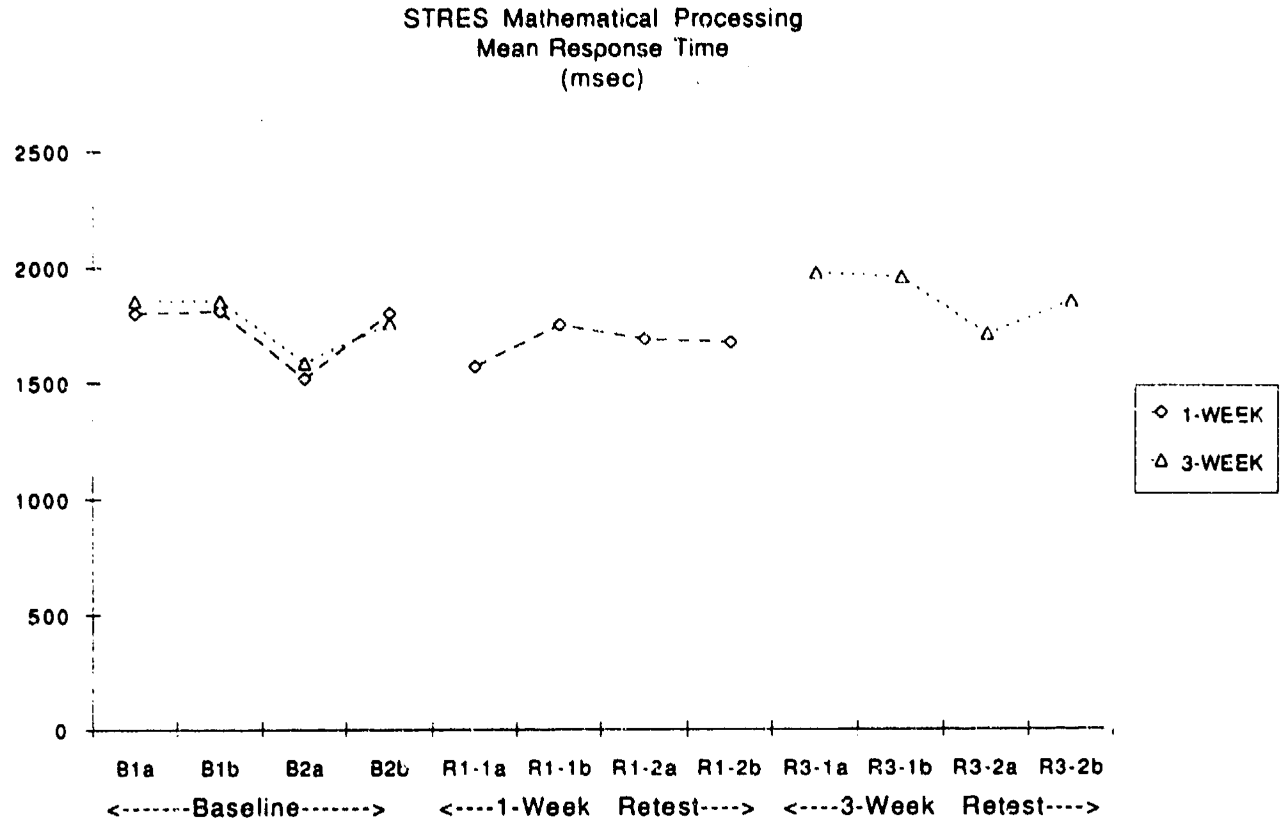

WRAIR Interval Production
Standard Deviation
(msec)
140
120
100
80
60
40
20
0
T1
T2
T3
T4
T5
B1
B2
<---------------------Training---------------------->
<----Baseline---->
75%ile
Mean
Median
25%ile

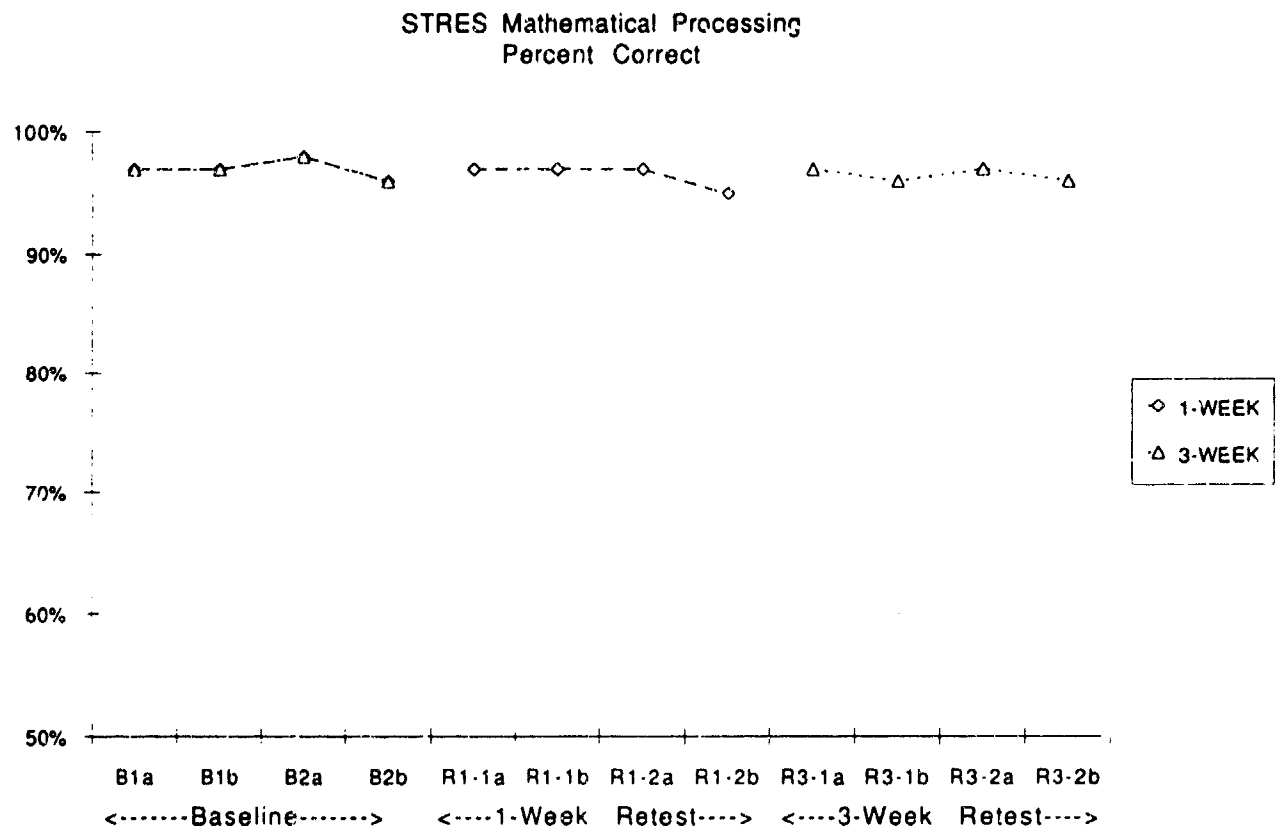
APPENDIX F

ONE-WEEK AND THREE-WEEK RETEST DATA

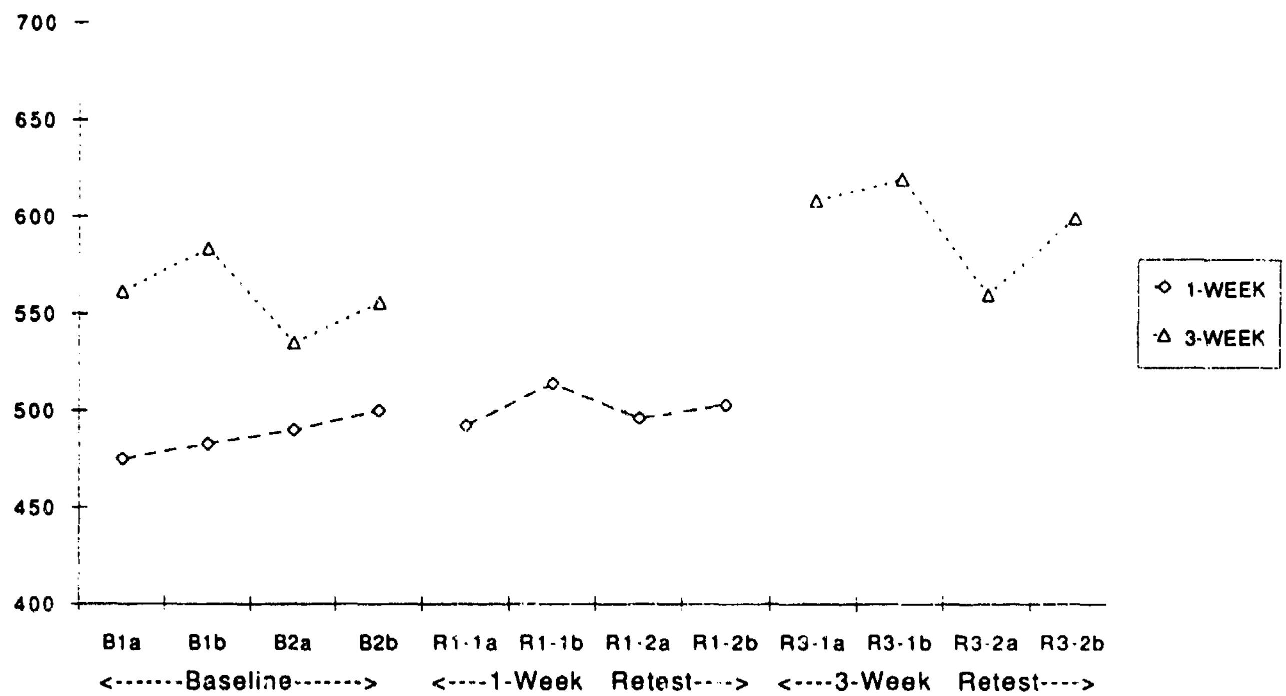

STRES Grammatical Reasoning
Mean Response Time
(msec)
6000
5500
5000
4500
4000
3500
3000
B1a
B1b
B2a
B2b
R1-1a
R1-1b
R1-2a
R1-2b
R3-1a
R3-1b
R3-2a
R3-2b
<-------Baseline------->
<----1-Week Retest---->
<----3-Week Retest---->
1-WEEK
3-WEEK


STRES Grammatical Reasoning
Percent Correct
100%
90%
80%
70%
60%
50%
B1a
B1b
B2a
B2b
R1-1a
R1-1b
R1-2a
R1-2b
R3-1a
R3-1b
R3-2a
R3-2b
<-------Baseline------->
1-Week Retest---->
1-WEEK
3-WEEK

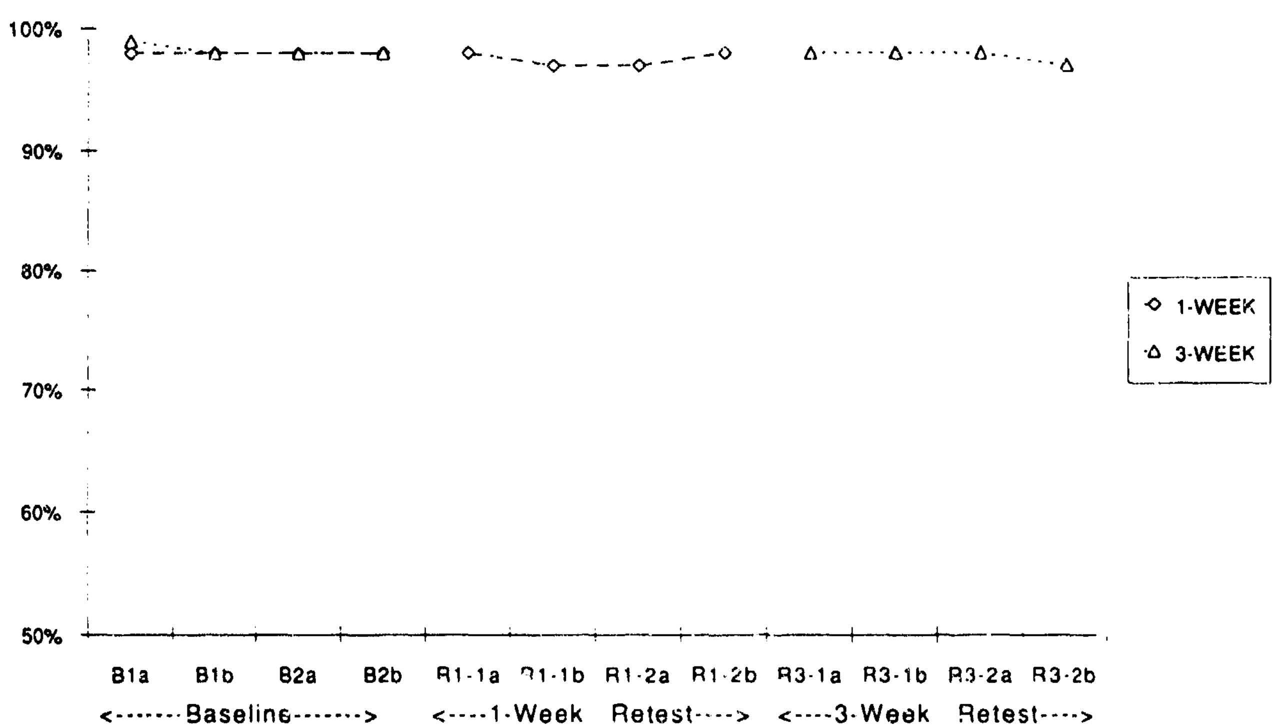

STRES Mathematical Processing
Mean Response Time
(msec)
2500
2000
1500
1000
500
0
B1a
B1b
B2a
B2b
R1-1a
R1-1b
R1-2a
R1-2b
R3-1a
R3-1b
R3-2a
R3-2b
<-------Baseline------->
<----1-Week Retest---->
<----3-Week Retest---->
1-WEEK
3-WEEK

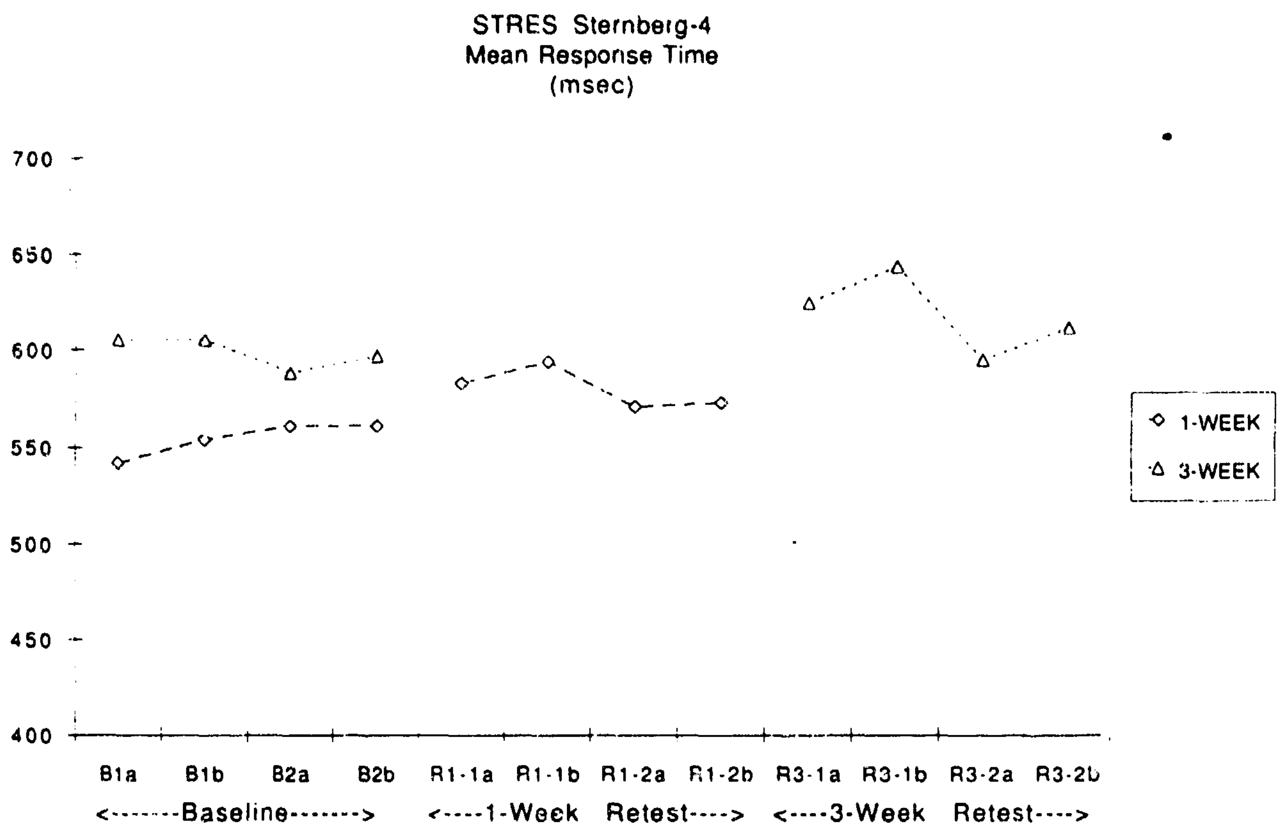

STRES Mathematical Processing
Percent Correct
100%
90%
80%
70%
60%
50%
B1a
B1b
B2a
B2b
R1-1a
R1-1b
R1-2a
R1-2b
R3-1a
R3-1b
R3-2a
R3-2b
<-------Baseline------->
<----1-Week Retest---->
<----3-Week Retest---->
1-WEEK
3-WEEK

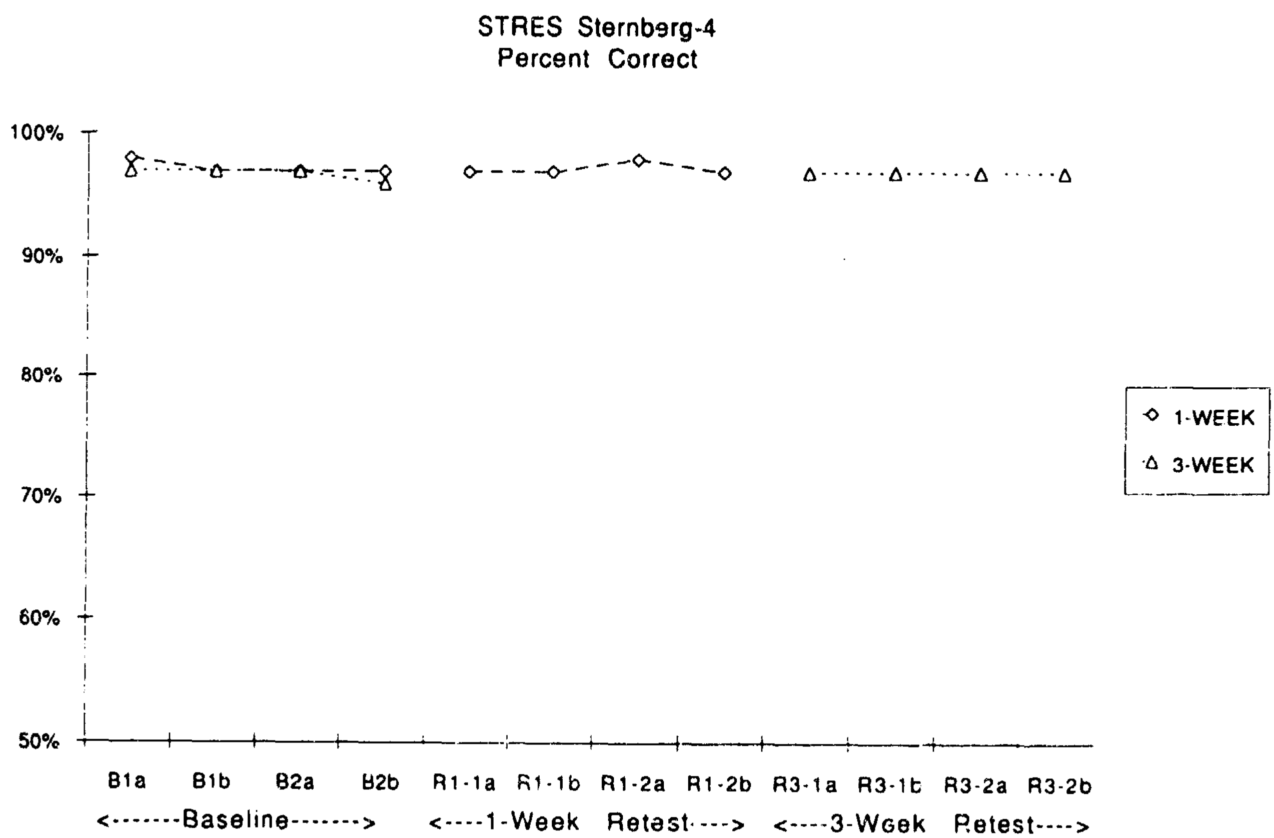
STRES Sternberg-2
Mean Response Time
(msec)



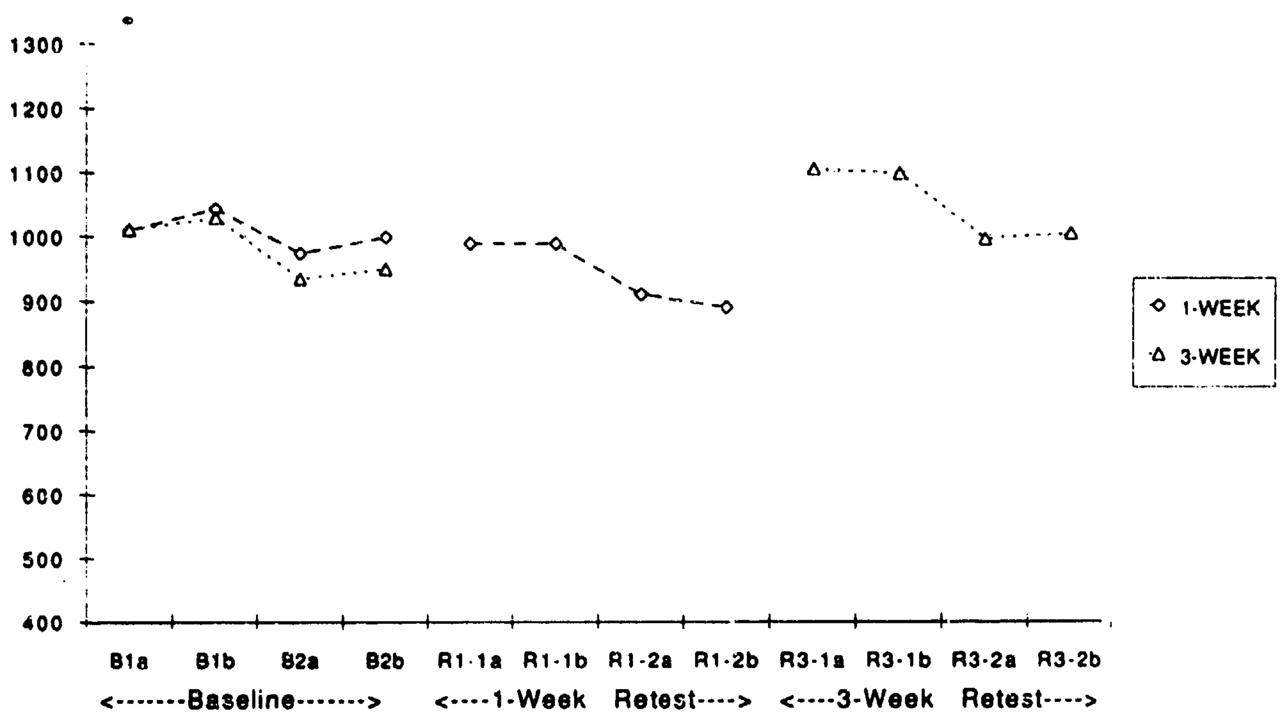
STRES Sternberg-2
Percent Correct



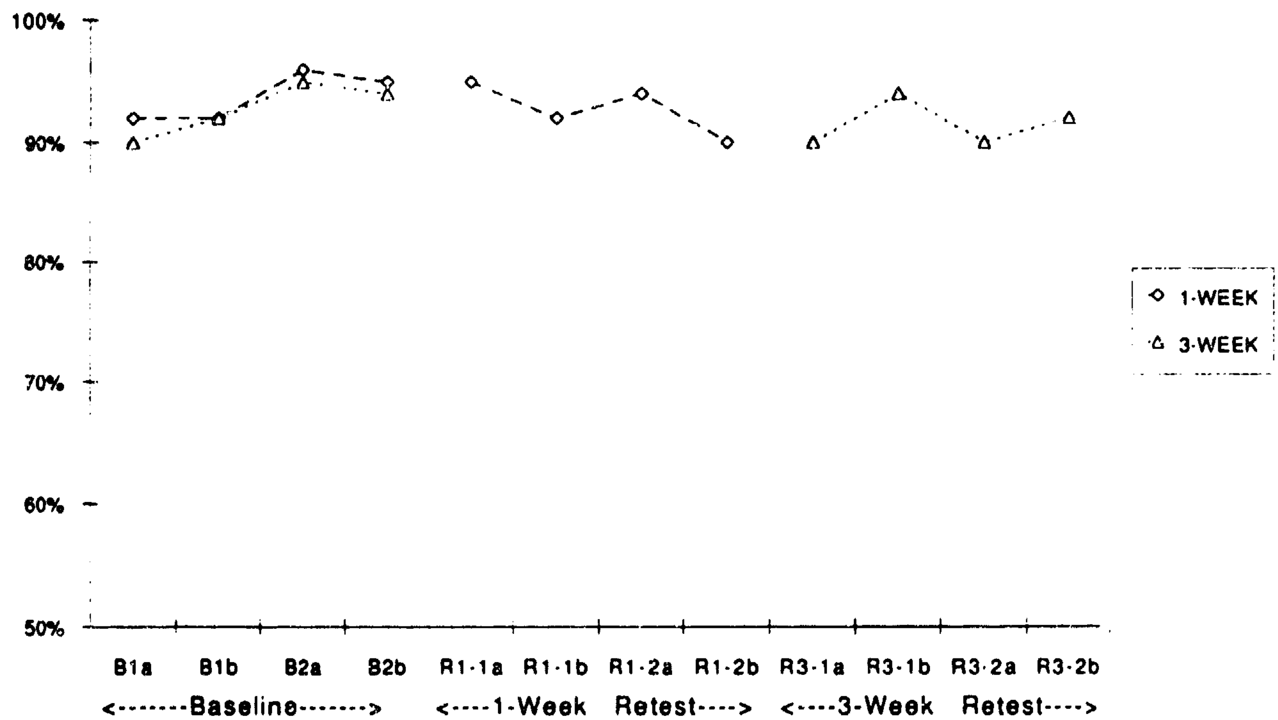

STRES Sternberg-4
Mean Response Time
(msec)
700
650
600
550
500
450
400
B1a
B1b
B2a
B2b
R1-1a
R1-1b
R1-2a
R1-2b
R3-1a
R3-1b
R3-2a
R3-2b
<-------Baseline------->
<----1-Week Retest---->
<----3-Week Retest---->
1-WEEK
3-WEEK

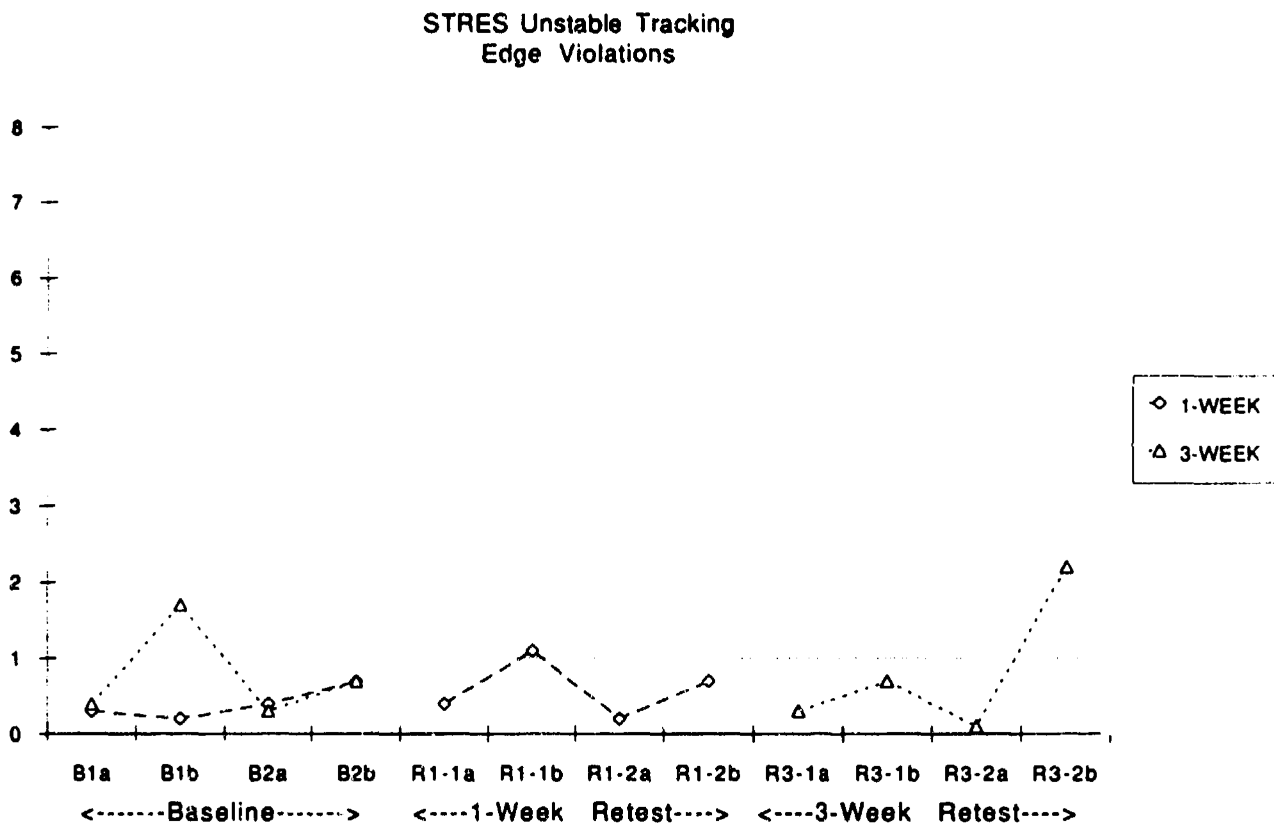

STRES Sternberg-4
Percent Correct
100%
90%
80%
70%
60%
50%
B1a
B1b
B2a
B2b
R1-1a
R1-1b
R1-2a
R1-2b
R3-1a
R3-1b
R3-2a
R3-2b
<-------Baseline------->
<----1-Week Retest---->
<----3-Week Retest---->
1-WEEK
3-WEEK

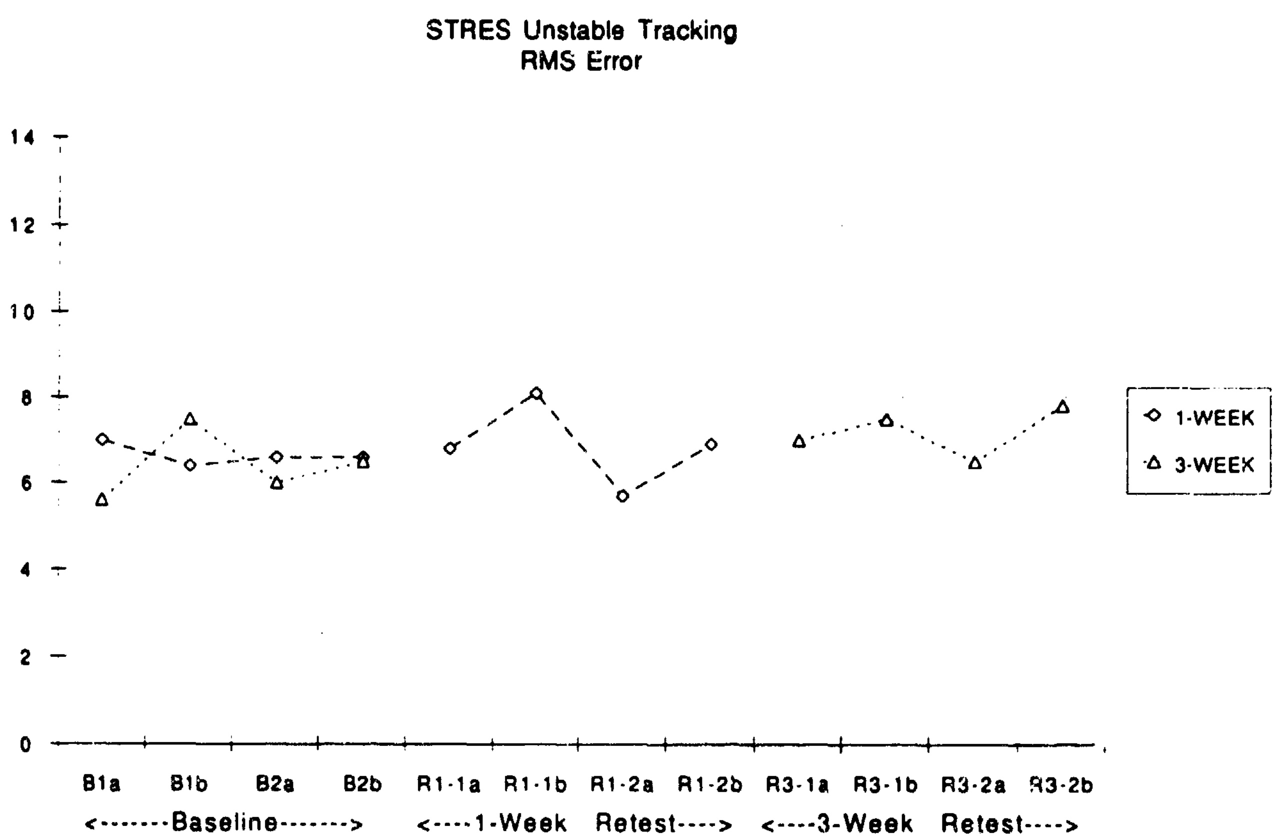
STRES Spatial Processing
Mean Response Time
(msec)



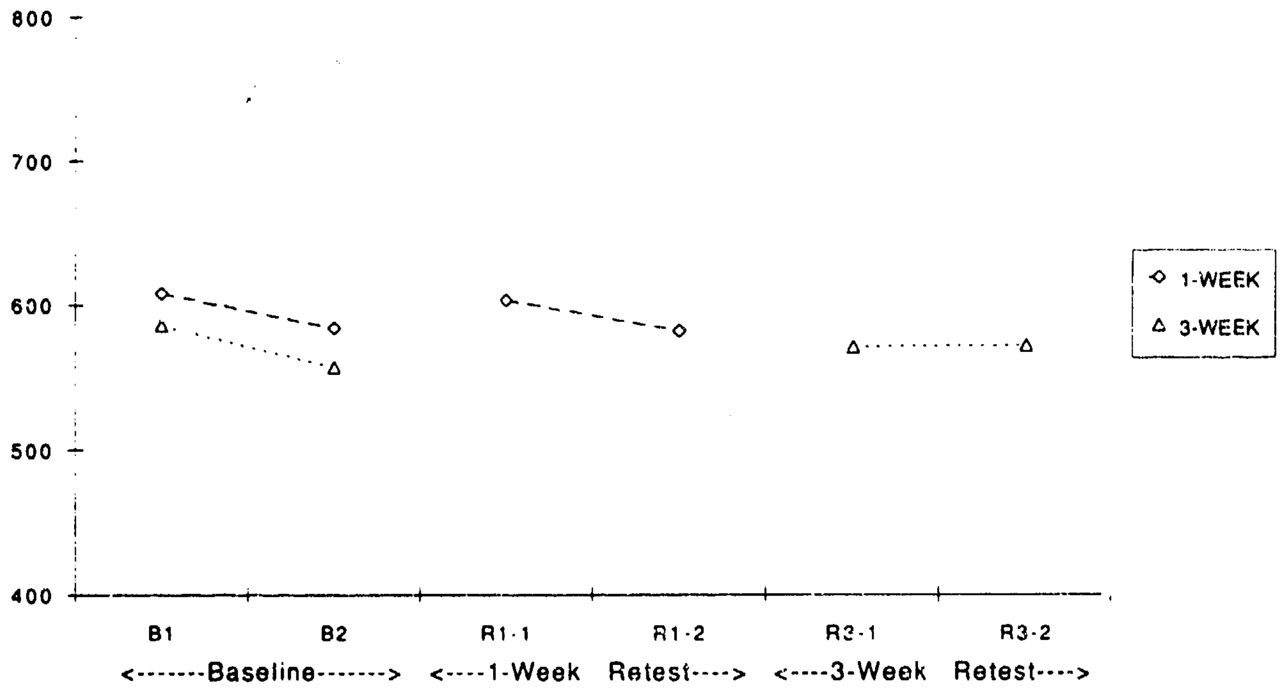
STRES Spatial Processing
Percent Correct



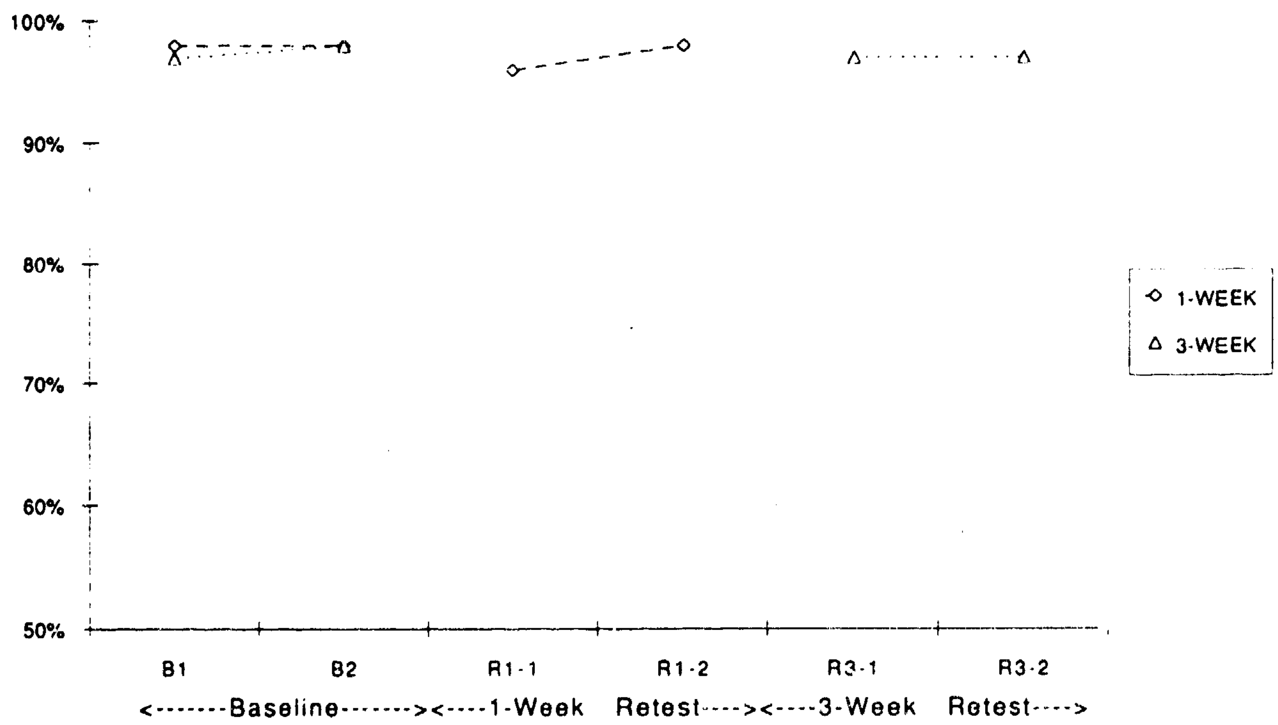

STRES Unstable Tracking
Edge Violations
8
7
6
5
4
3
2
1
0
B1a
B1b
B2a
B2b
R1-1a
R1-1b
R1-2a
R1-2b
R3-1a
R3-1b
R3-2a
R3-2b
<-------Baseline------->
<----1-Week Retest---->
<----3-Week Retest---->
1-WEEK
3-WEEK


STRES Unstable Tracking
RMS Error
14
12
10
8
6
4
2
0
B1a
B1b
B2a
B2b
R1-1a
R1-1b
R1-2a
R1-2b
R3-1a
R3-1b
R3-2a
R3-2b
<-------Baseline------->
<----1-Week Retest---->
<----3-Week Retest---->
1-WEEK
3-WEEK

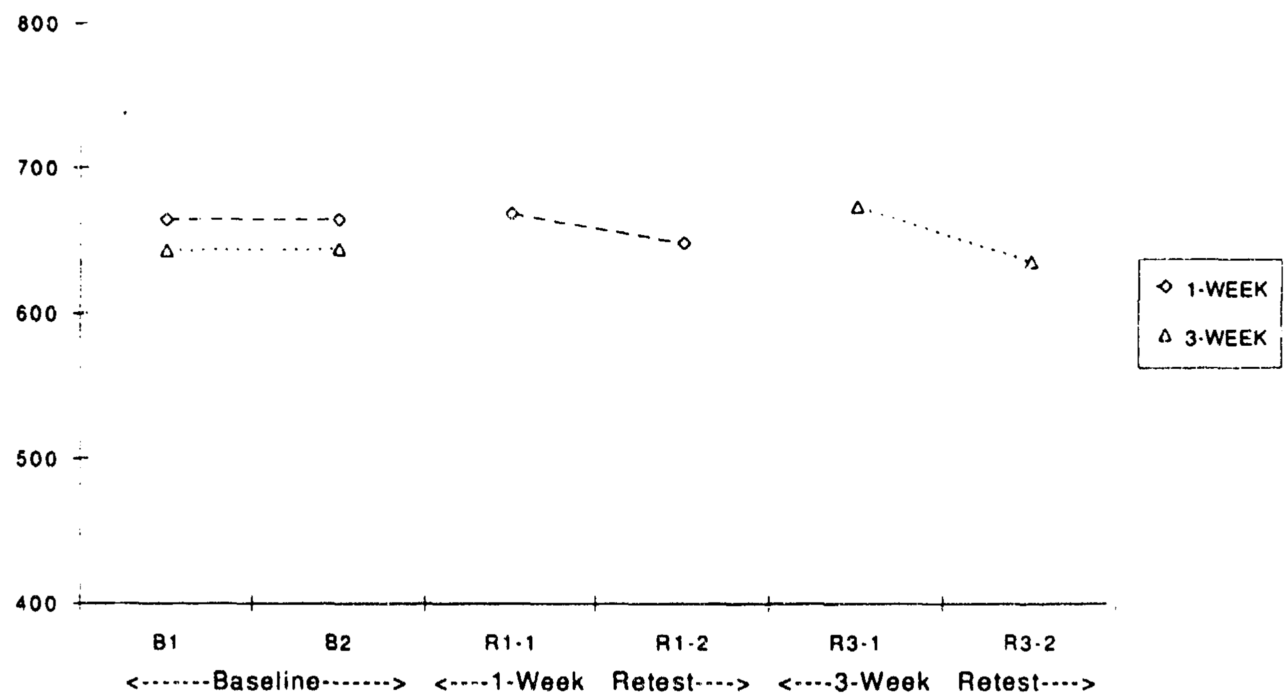
STRES Reaction Time - BASIC (1)
Mean Response Time
(msec)



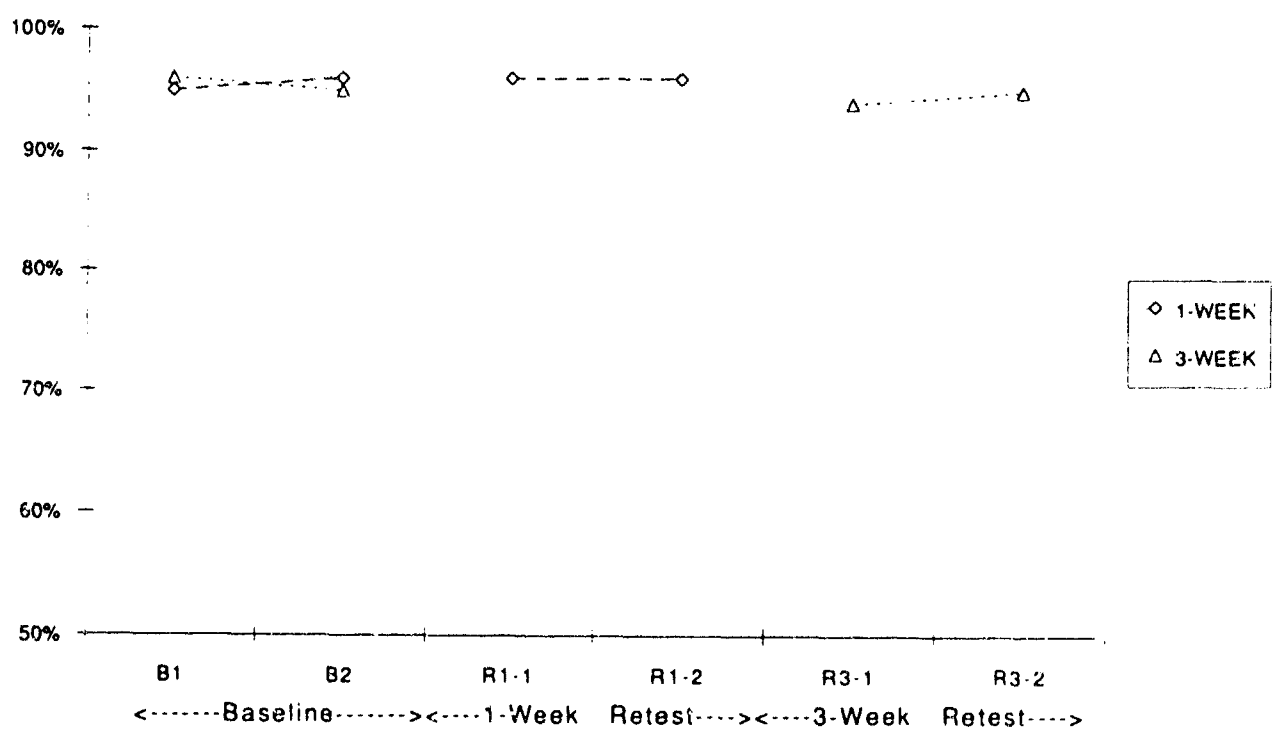
STRES Reaction Time - BASIC (1)
Percent Correct



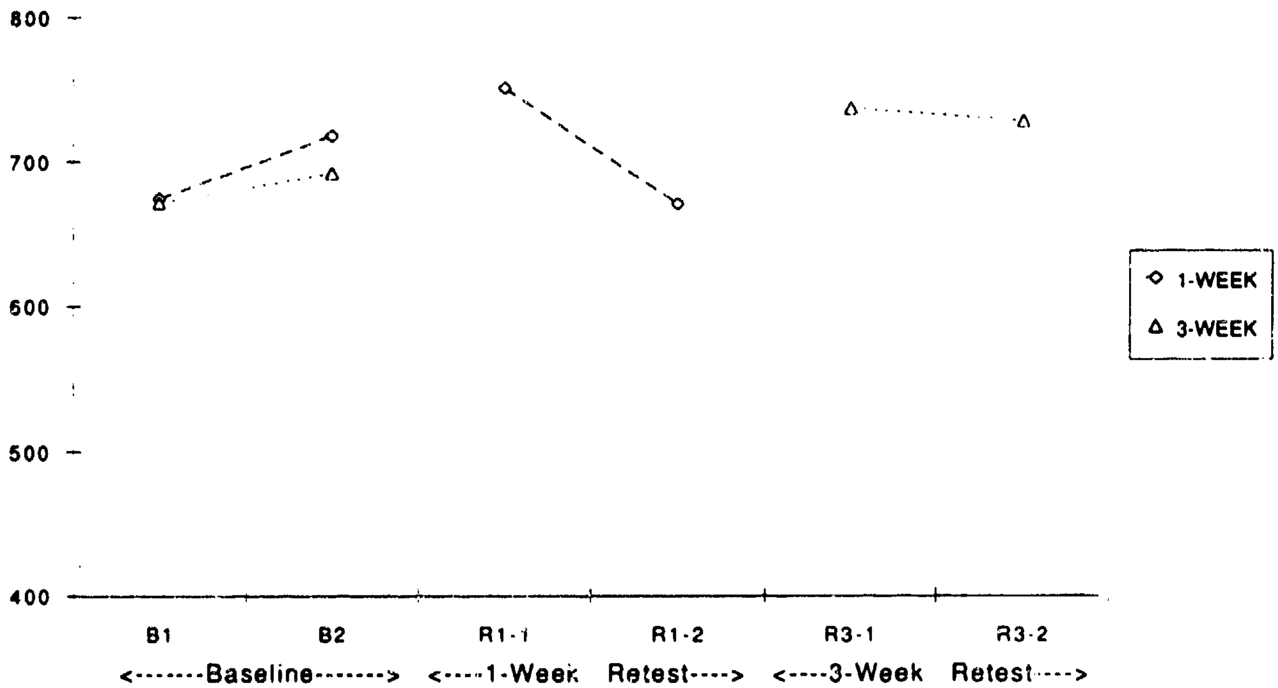
STRES Reaction Time - CODED (2)
Mean Response Time
(msec)



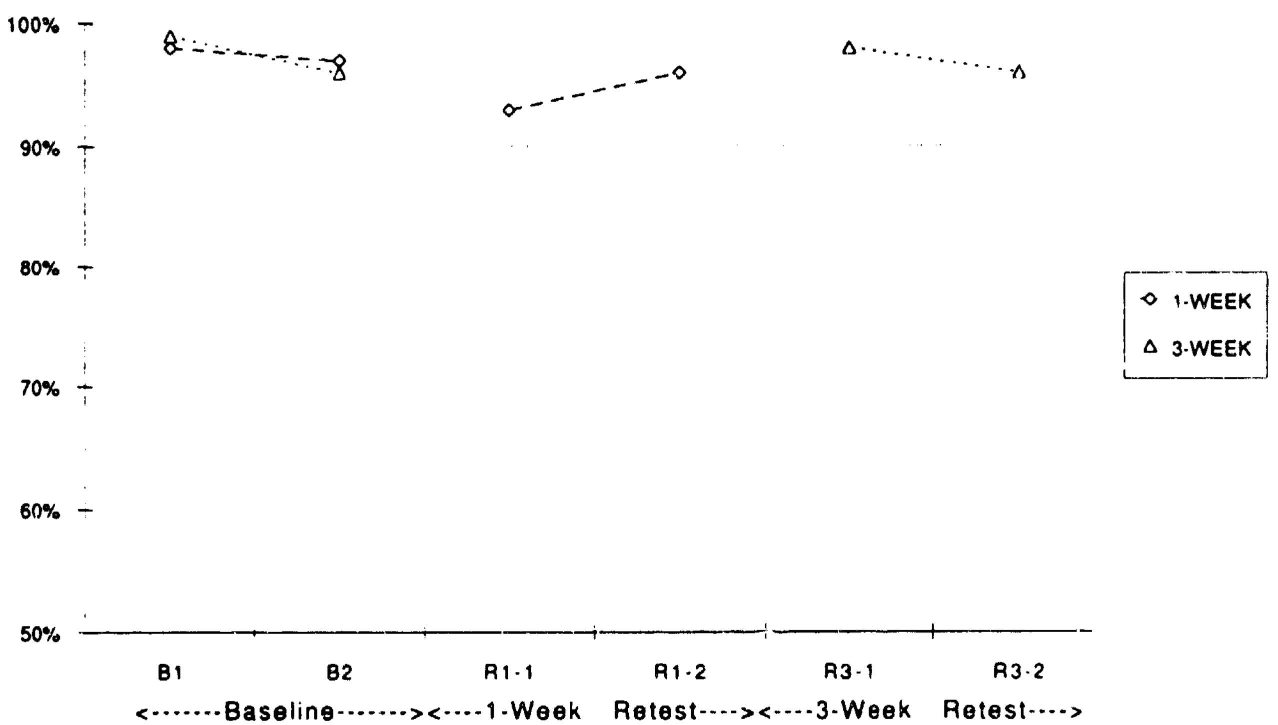
STRES Reaction Time - CODED (2)
Percent Correct

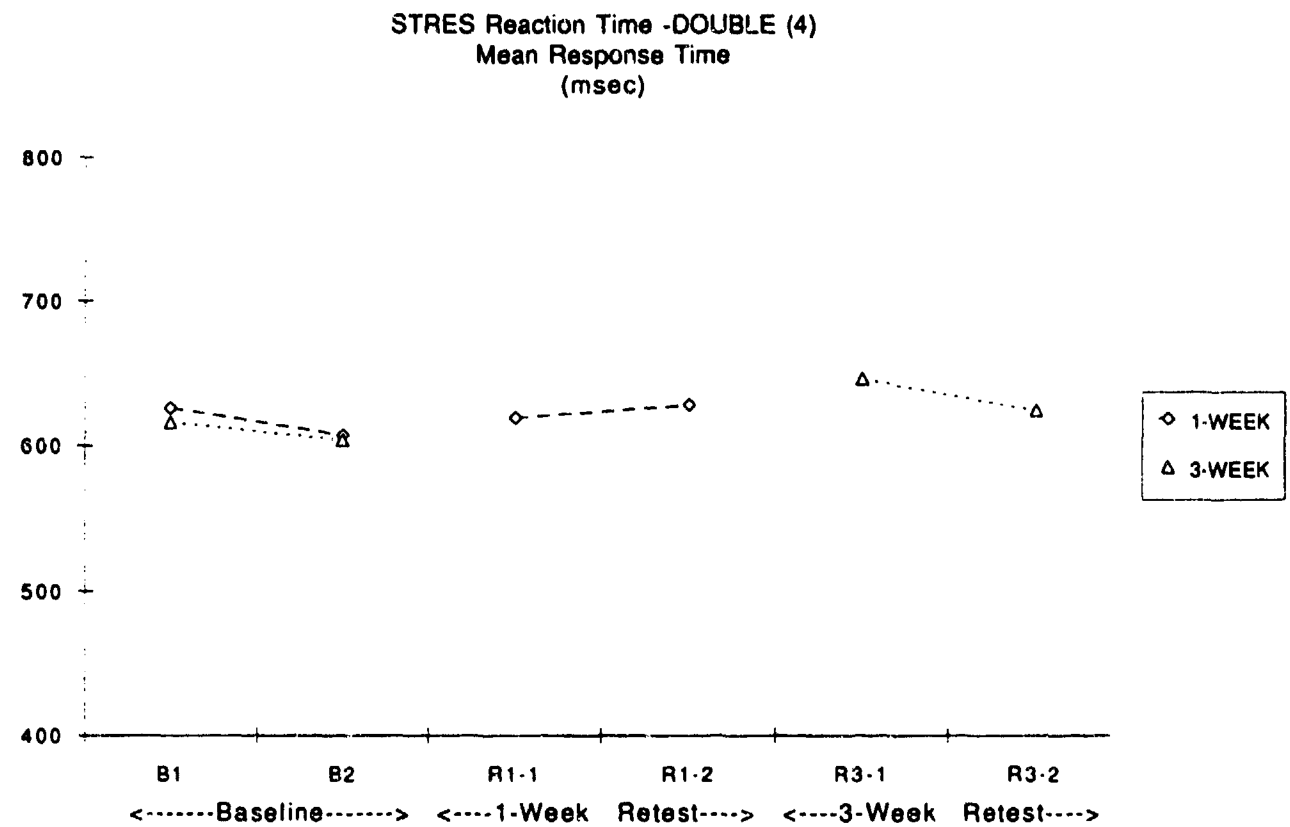


STRES Reaction Time -UNCERT (3)
Mean Response Time
(msec)

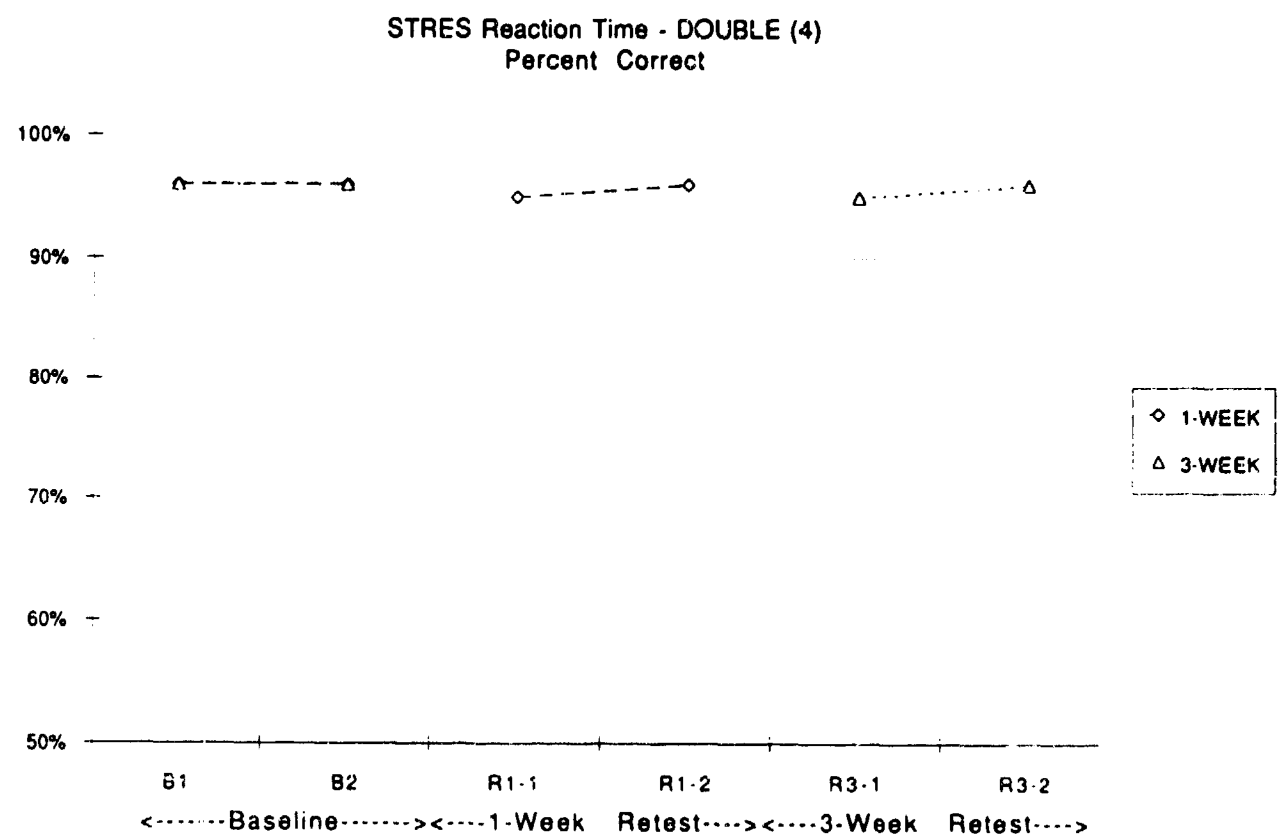


STRES Reaction Time - UNCERT (3)
Percent Correct

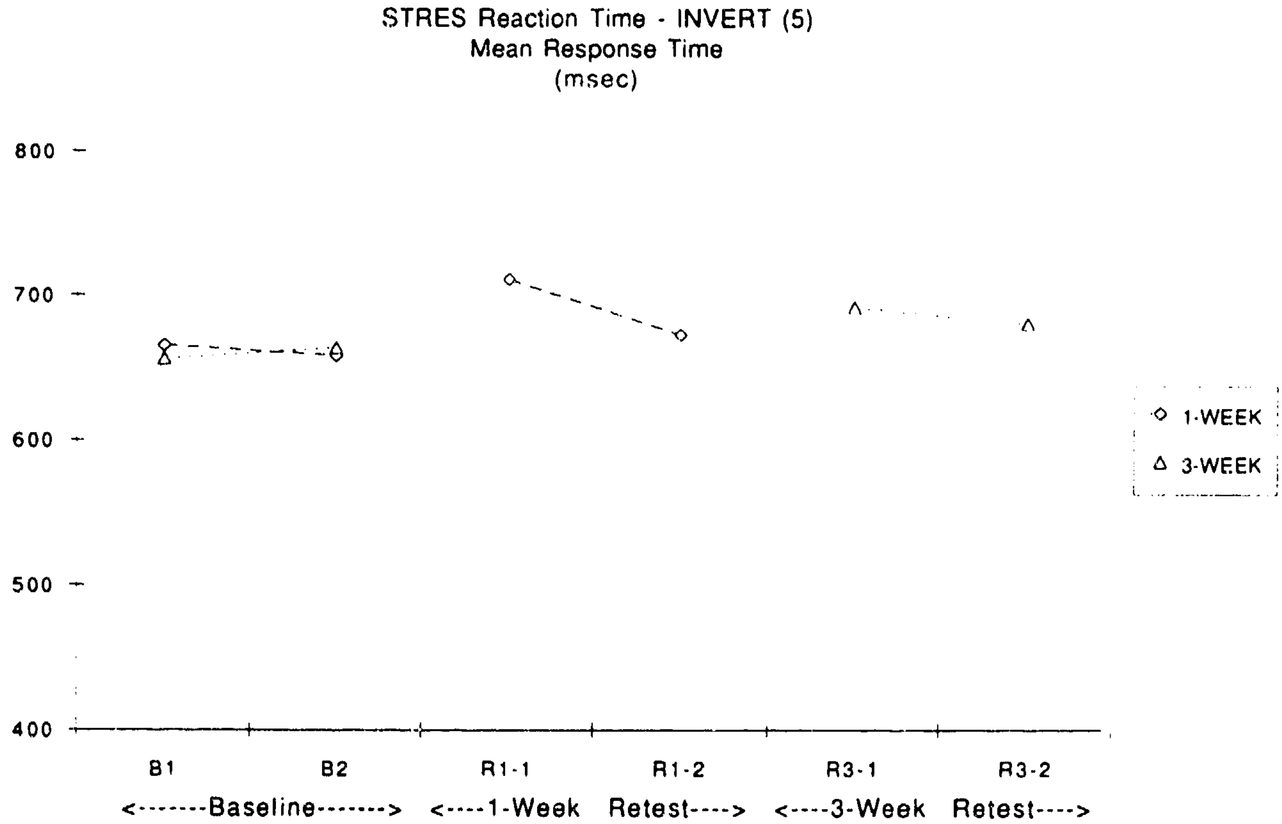


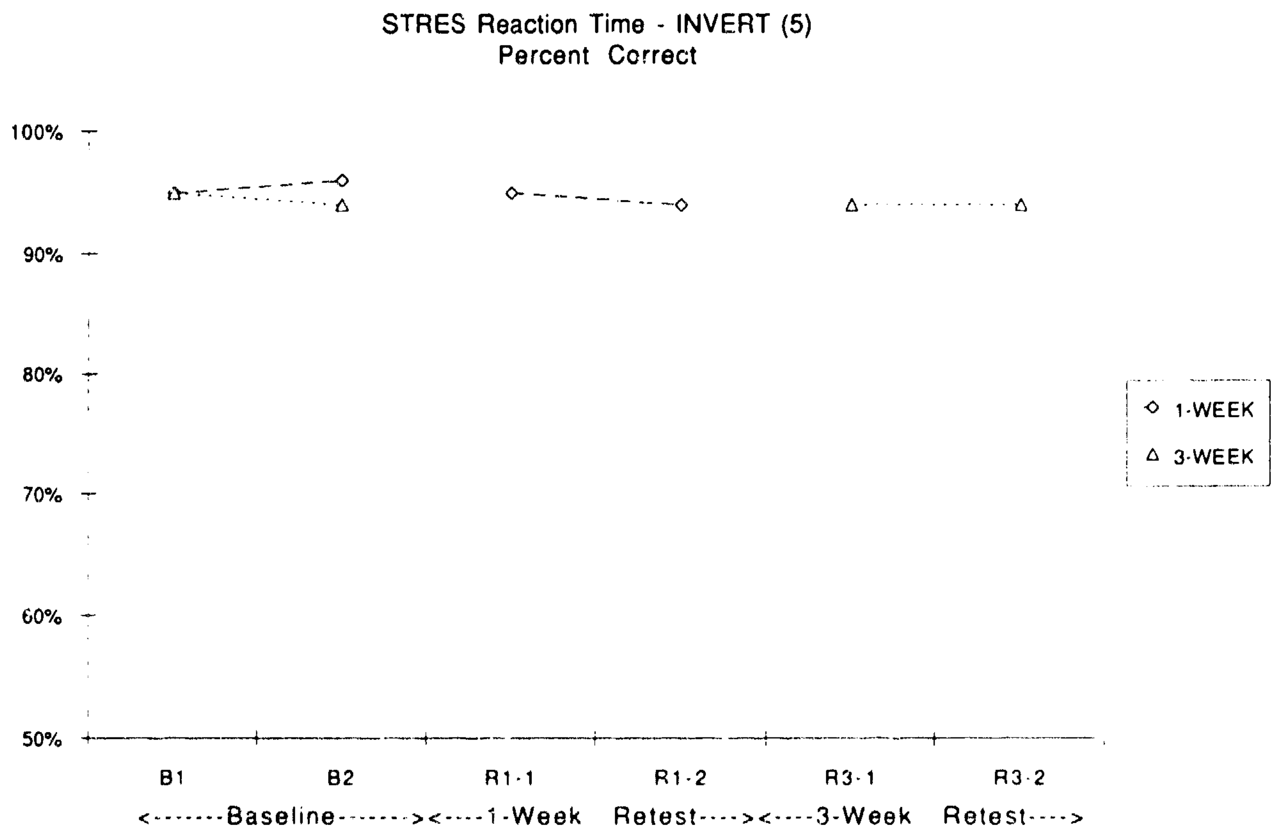


STRES Reaction Time -DOUBLE (4)
Mean Response Time
(msec)
800
700
600
500
400
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------> <----1-Week Retest----> <----3-Week Retest---->
1-WEEK
3-WEEK

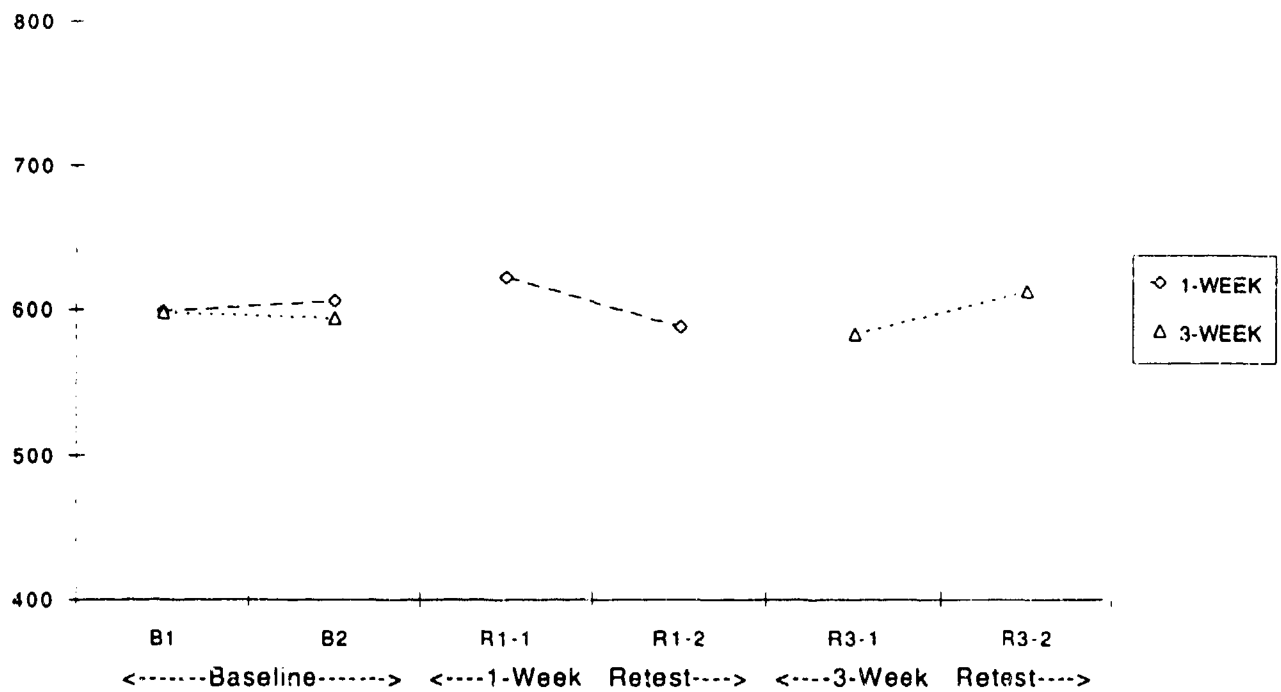


STRES Reaction Time - DOUBLE (4)
Percent Correct
100%
90%
80%
70%
60%
50%
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------><----1-Week Retest----><----3-Week Retest---->
1-WEEK
3-WEEK

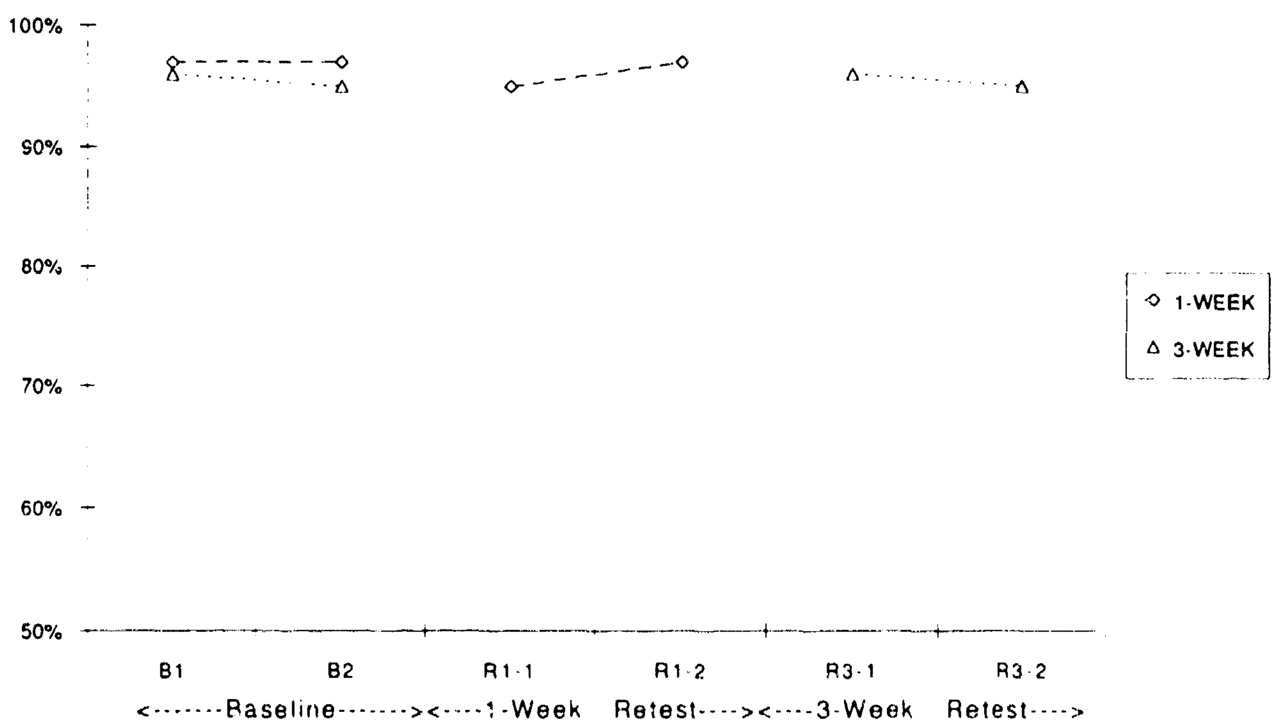

STRES Reaction Time - INVERT (5)
Mean Response Time
(msec)
800
700
600
500
400
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------> <----1-Week Retest----> <----3-Week Retest---->
1-WEEK
3-WEEK

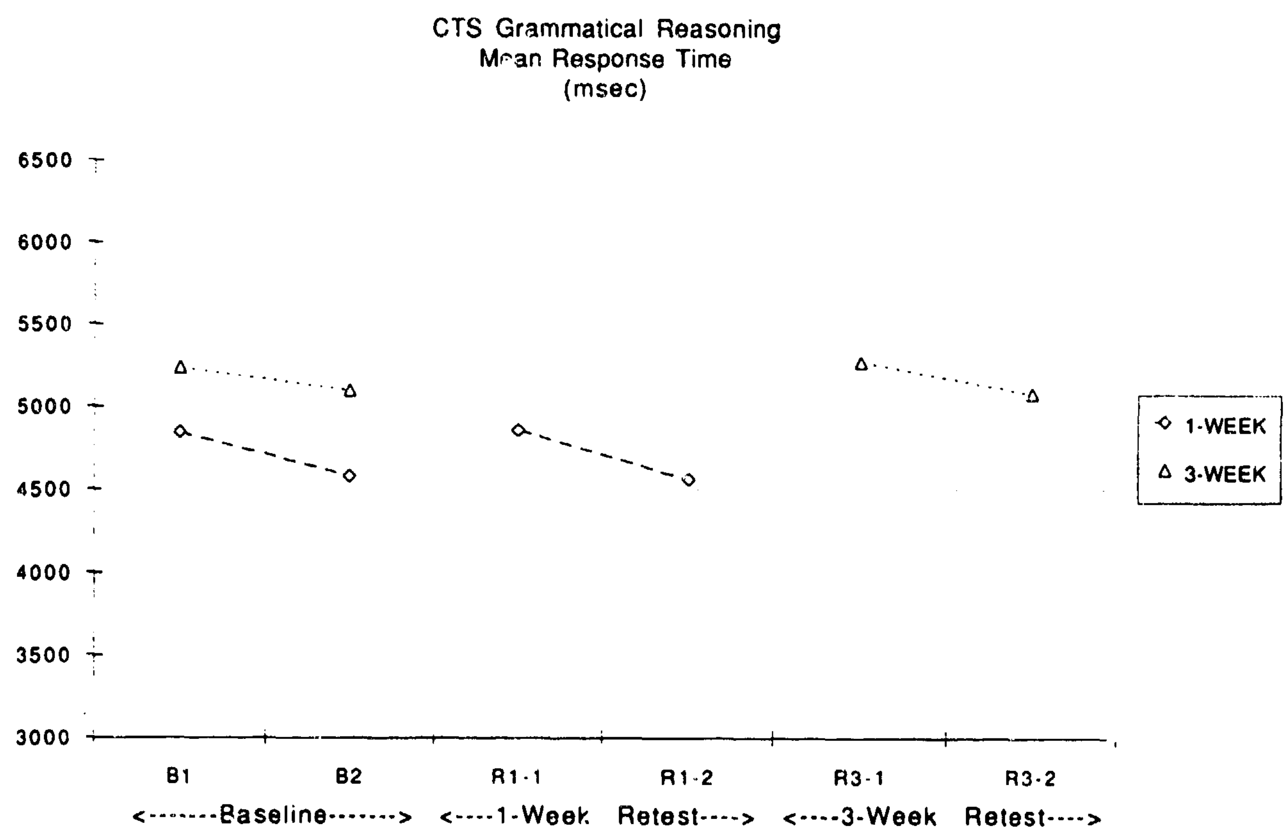

STRES Reaction Time - INVERT (5)
Percent Correct
100%
90%
80%
70%
60%
50%
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------><----1-Week Retest----><----3-Week Retest---->
1-WEEK
3-WEEK

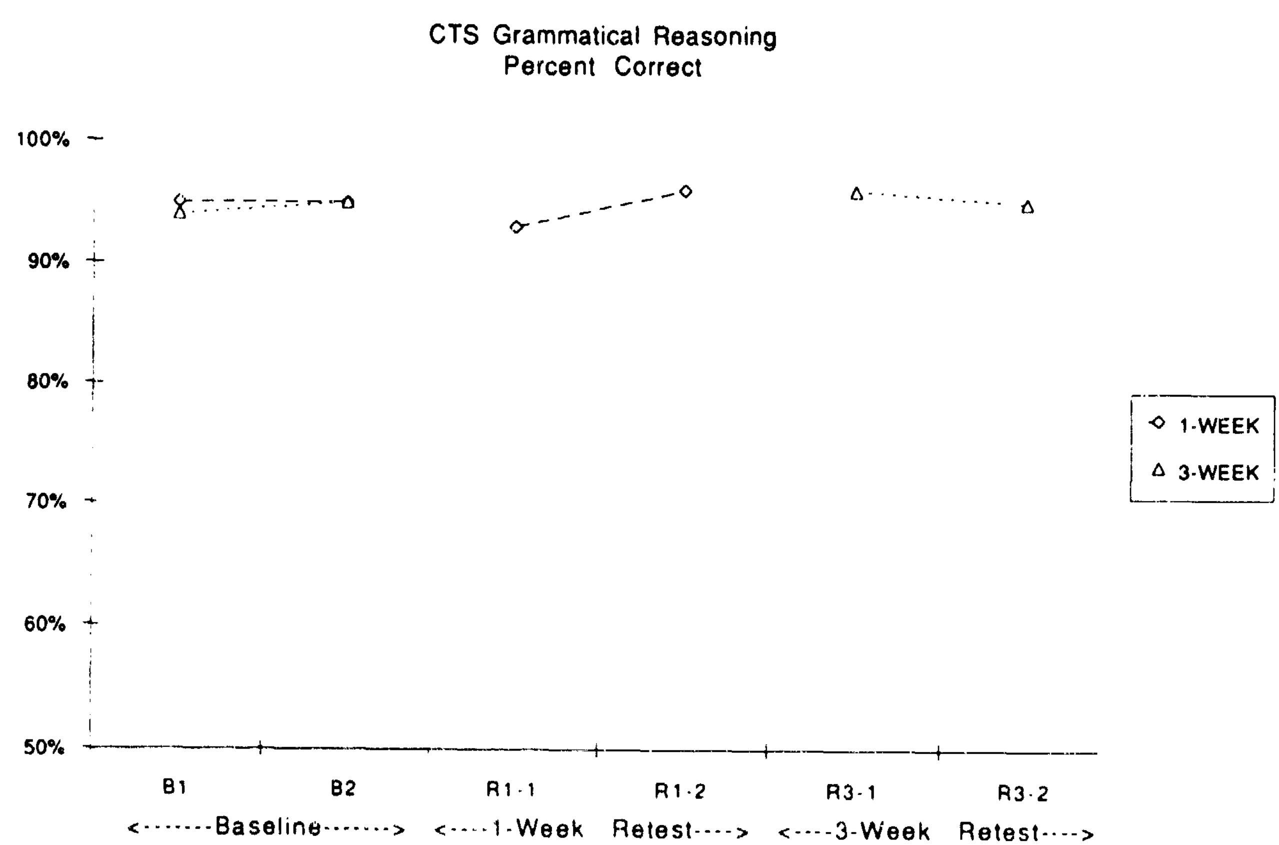
STRES Reaction Time - BASIC (6)
Mean Response Time
(msec)

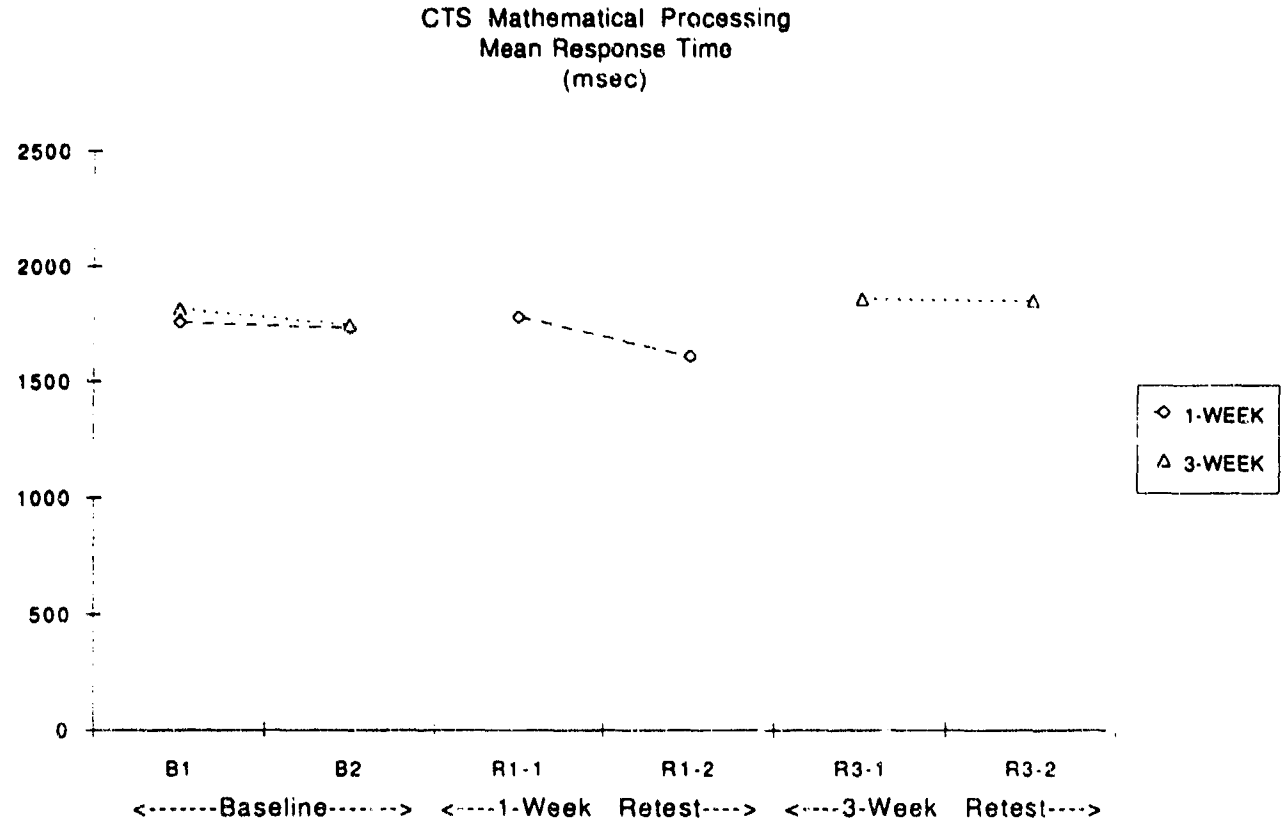


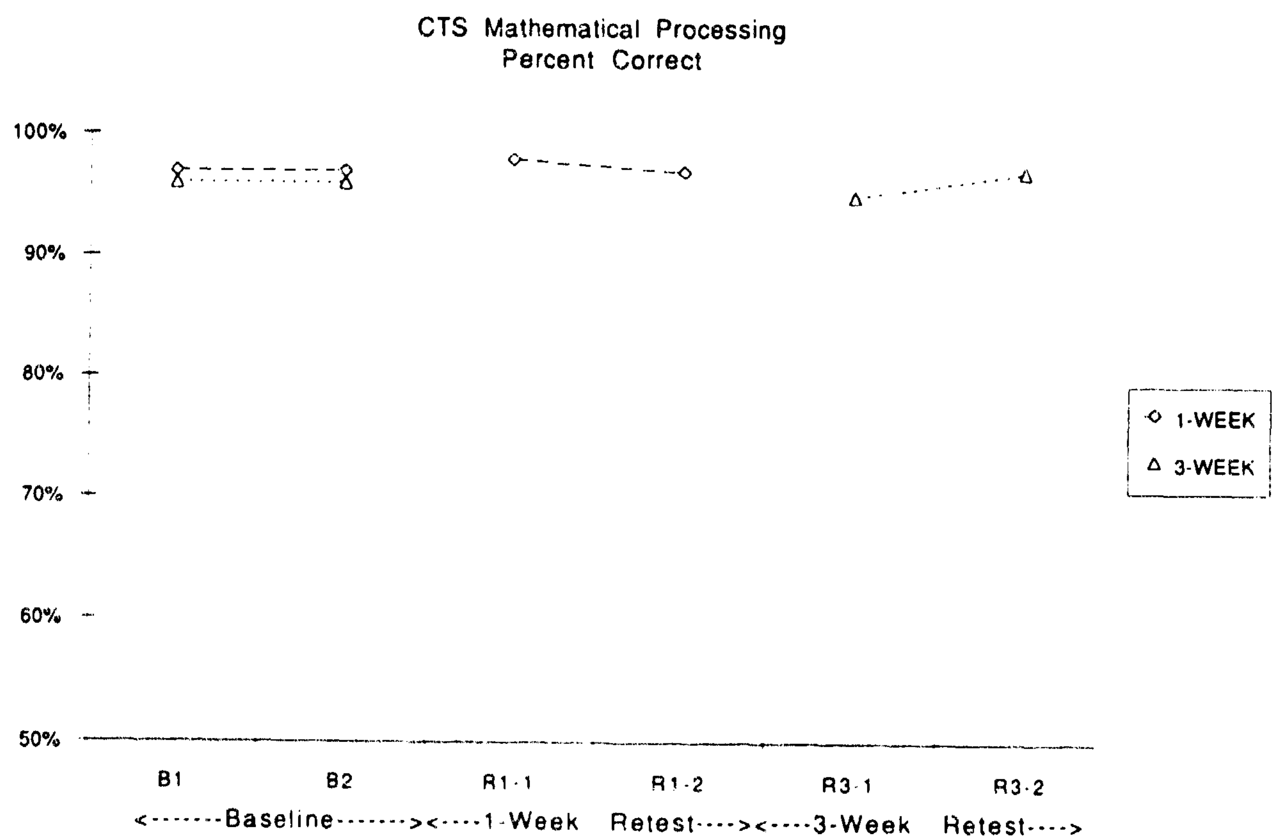
STRES Reaction Time - BASIC (6)
Percent Correct

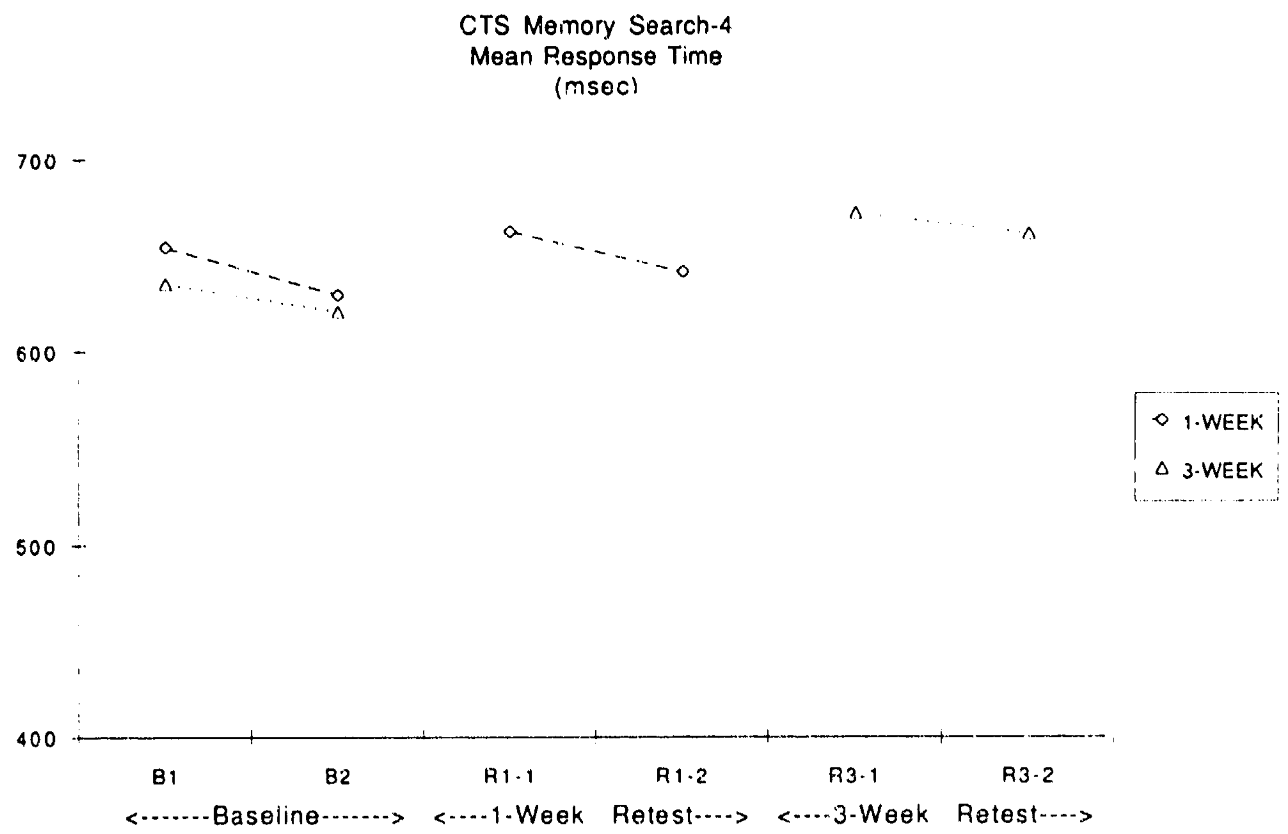


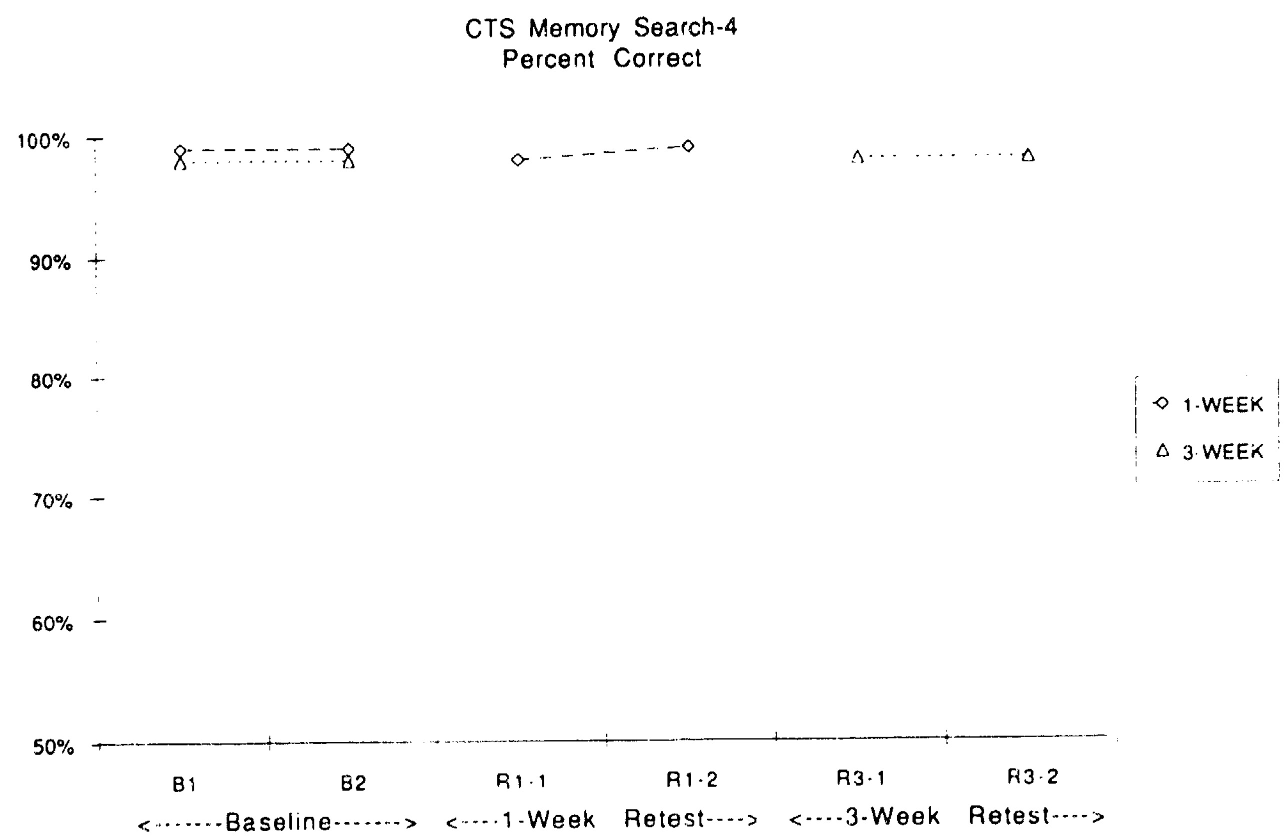

CTS Grammatical Reasoning
Mean Response Time
(msec)
6500
6000
5500
5000
4500
4000
3500
3000
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------> <----1-Week Retest----> <----3-Week Retest---->
1-WEEK
3-WEEK

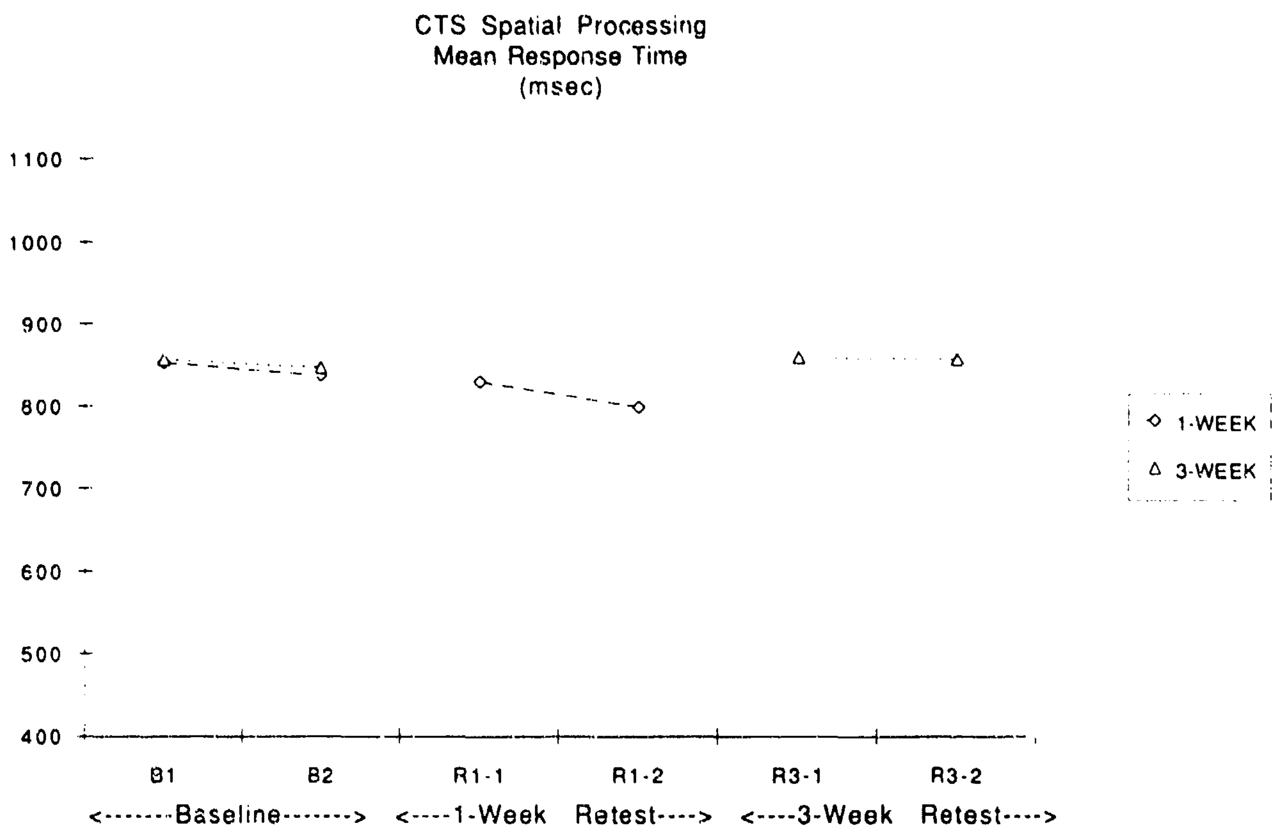

CTS Grammatical Reasoning
Percent Correct
100%
90%
80%
70%
60%
50%
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------> <----1-Week Retest----> <----3-Week Retest---->
1-WEEK
3-WEEK

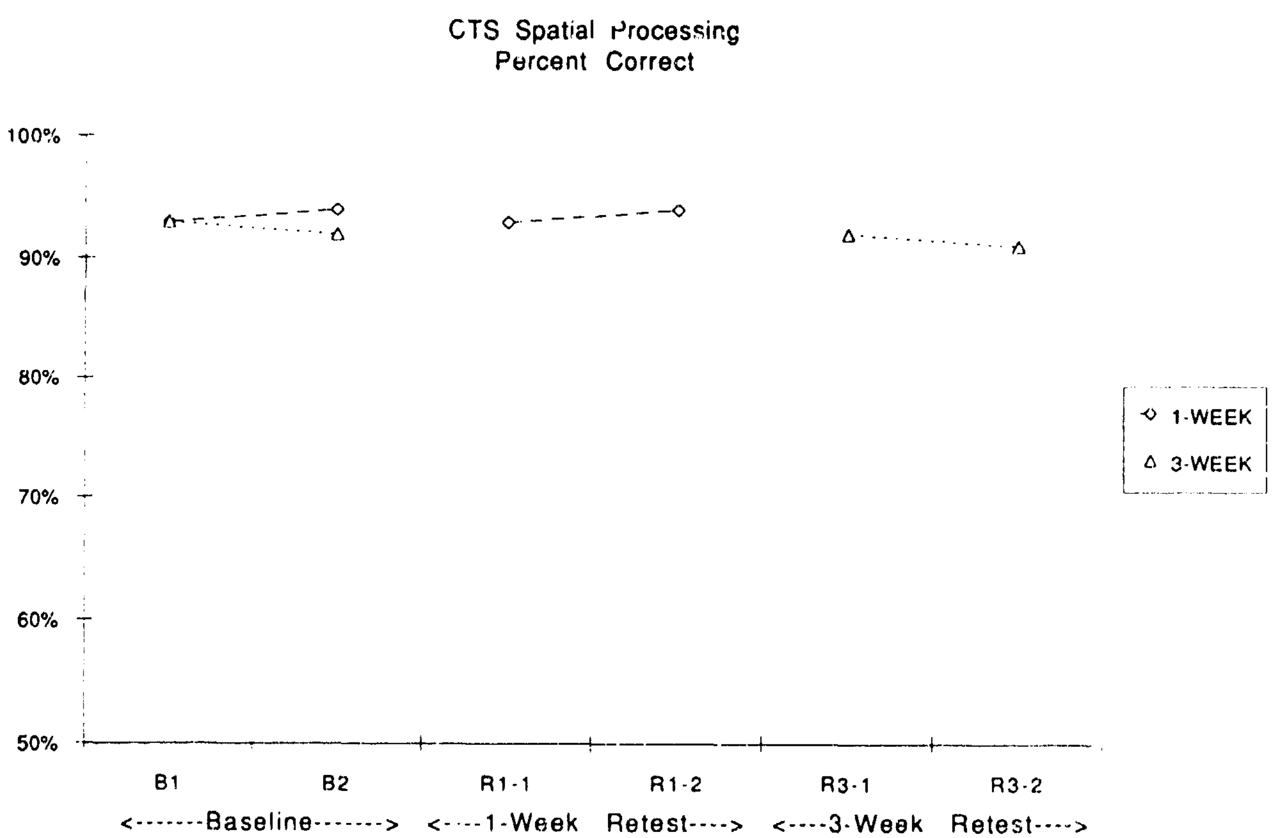

CTS Mathematical Processing
Mean Response Time
(msec)
2500
2000
1500
1000
500
0
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline------> <----1-Week Retest----> <----3-Week Retest---->
1-WEEK
3-WEEK


CTS Mathematical Processing
Percent Correct
100%
90%
80%
70%
60%
50%
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------><----1-Week Retest----><----3-Week Retest---->
1-WEEK
3-WEEK

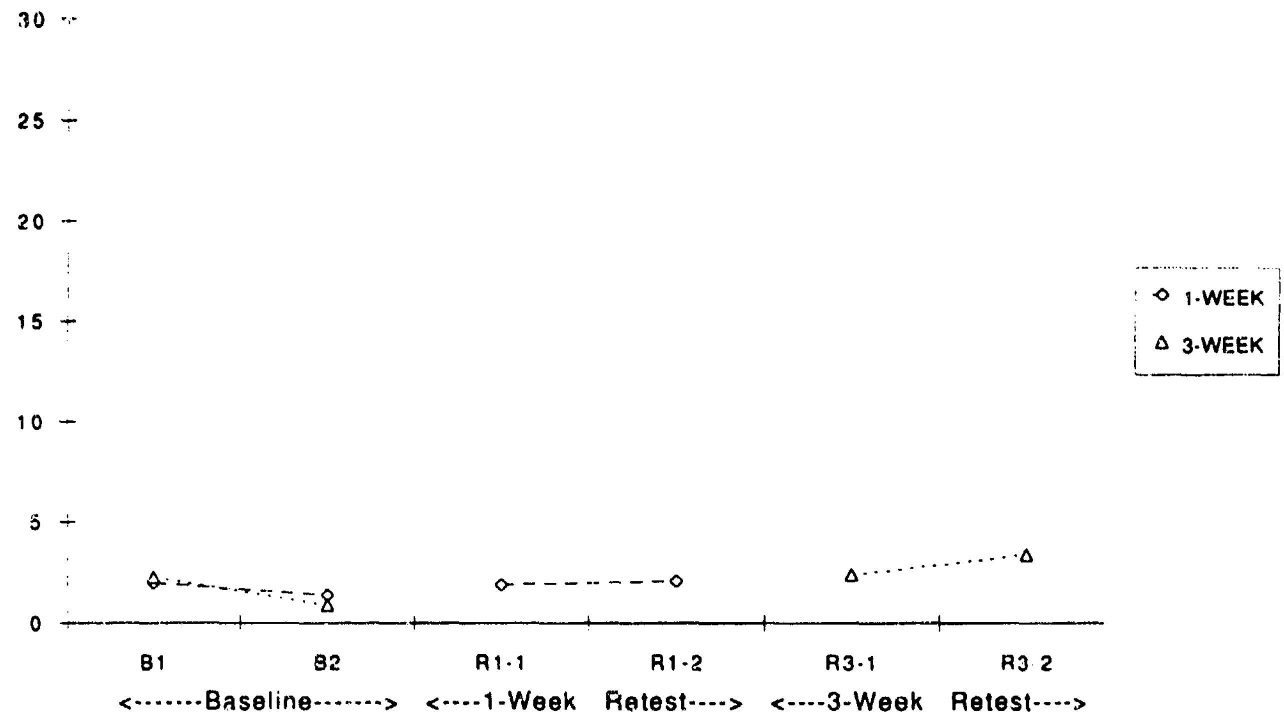

CTS Memory Search-4
Mean Response Time
(msec)
700
600
500
400
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------> <----1-Week Retest----> <----3-Week Retest---->
1-WEEK
3-WEEK


CTS Memory Search-4
Percent Correct
100%
90%
80%
70%
60%
50%
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------> <----1-Week Retest----> <----3-Week Retest---->
1-WEEK
3-WEEK

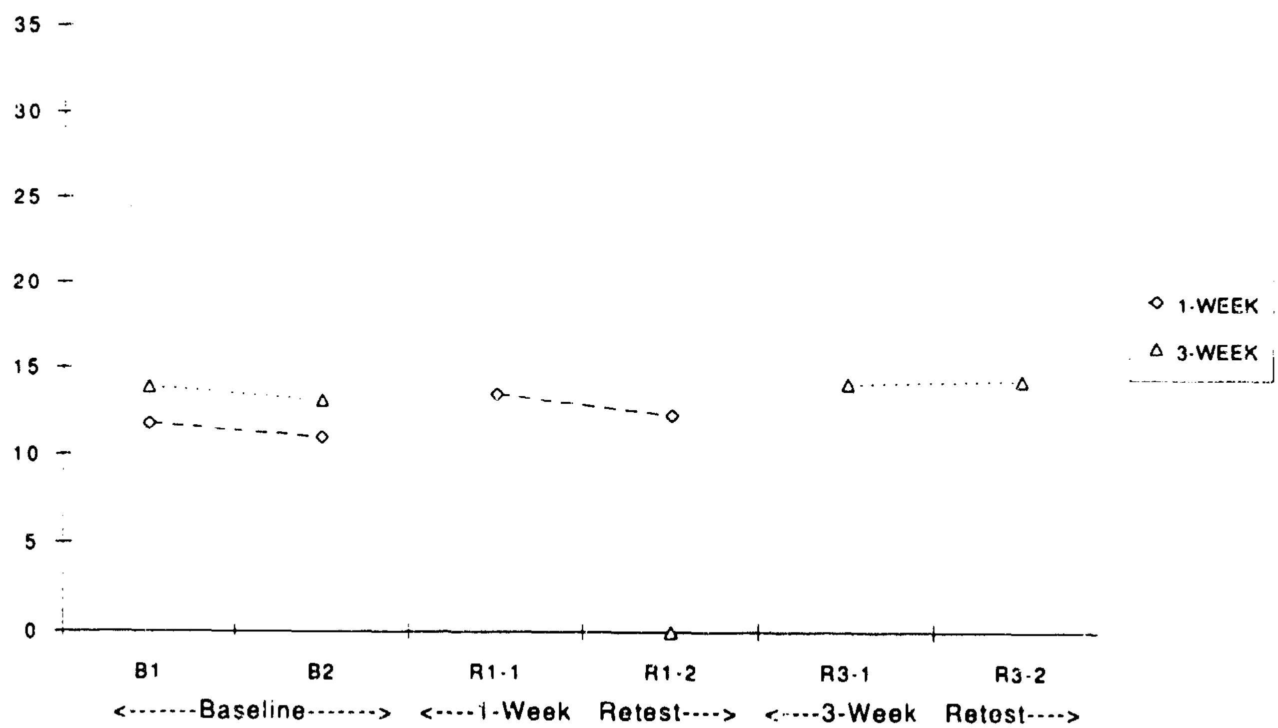

CTS Spatial Processing
Mean Response Time
(msec)
1100
1000
900
800
700
600
500
400
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------> <----1-Week Retest----> <----3-Week Retest---->
1-WEEK
3-WEEK


CTS Spatial Processing
Percent Correct
100%
90%
80%
70%
60%
50%
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------> <----1-Week Retest----> <----3-Week Retest---->
1-WEEK
3-WEEK

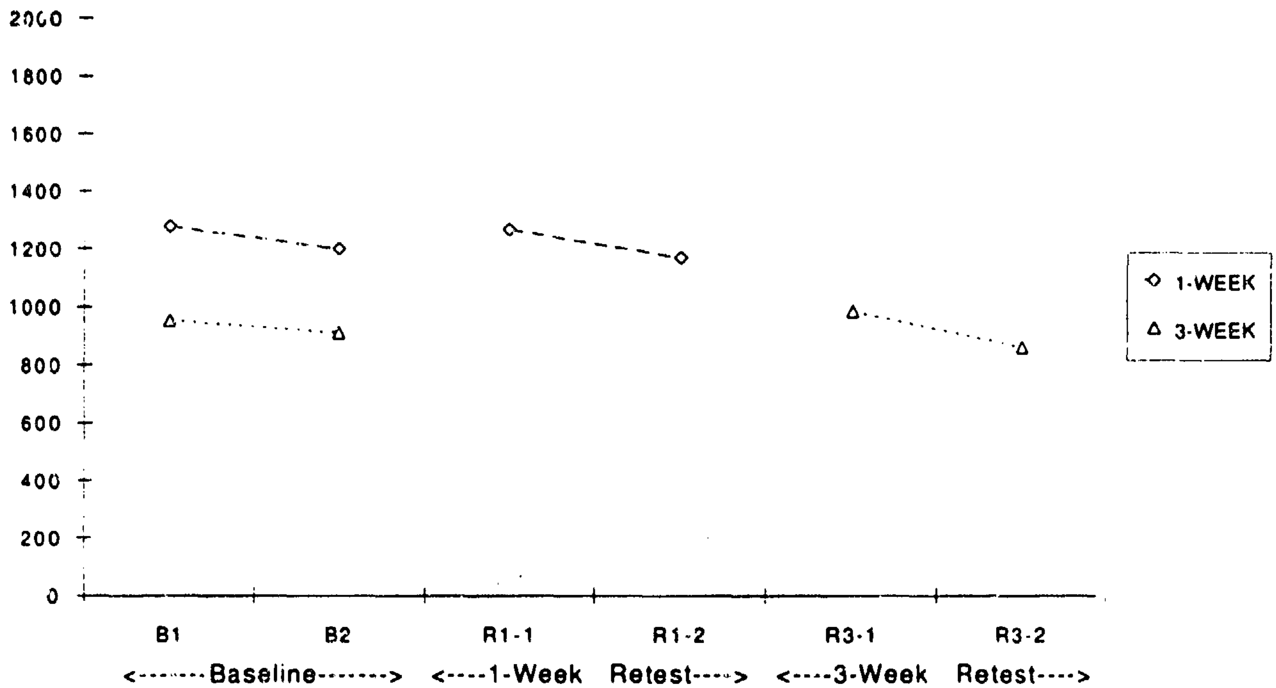
CTS Unstable Tracking
Edge Violations



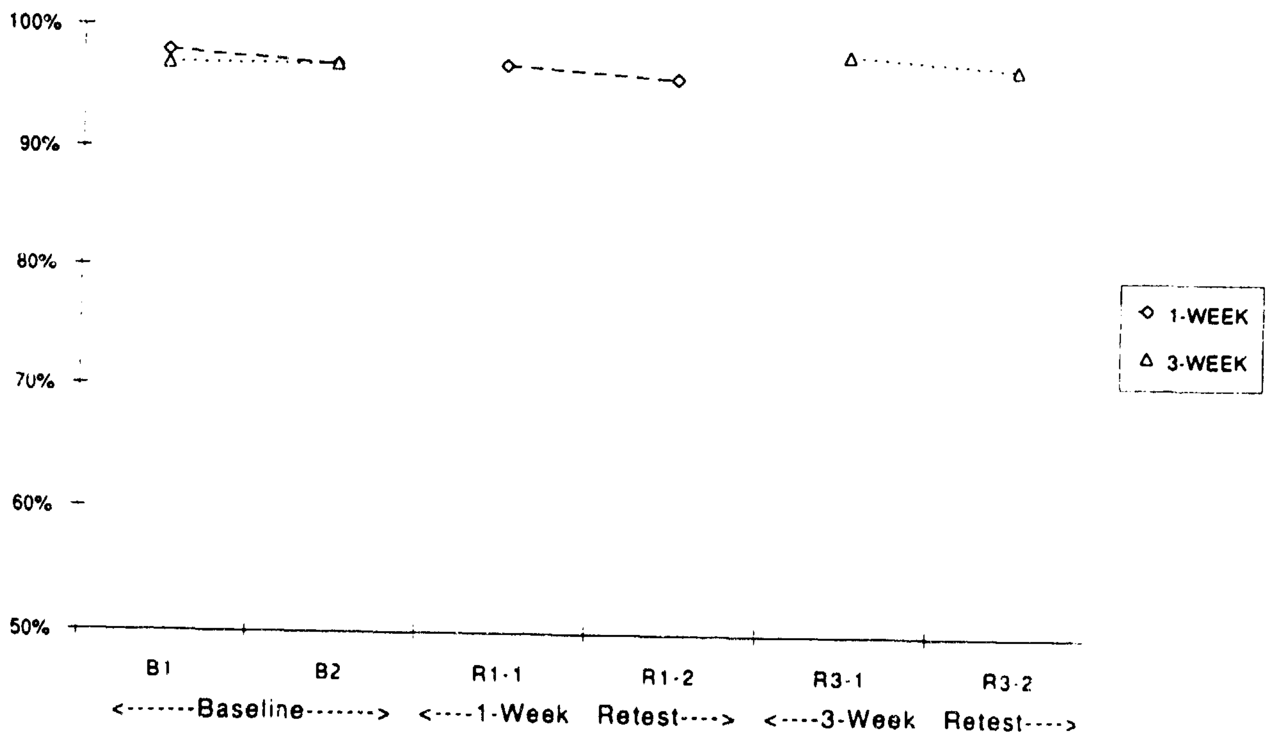
CTS Unstable Tracking
RMS Error

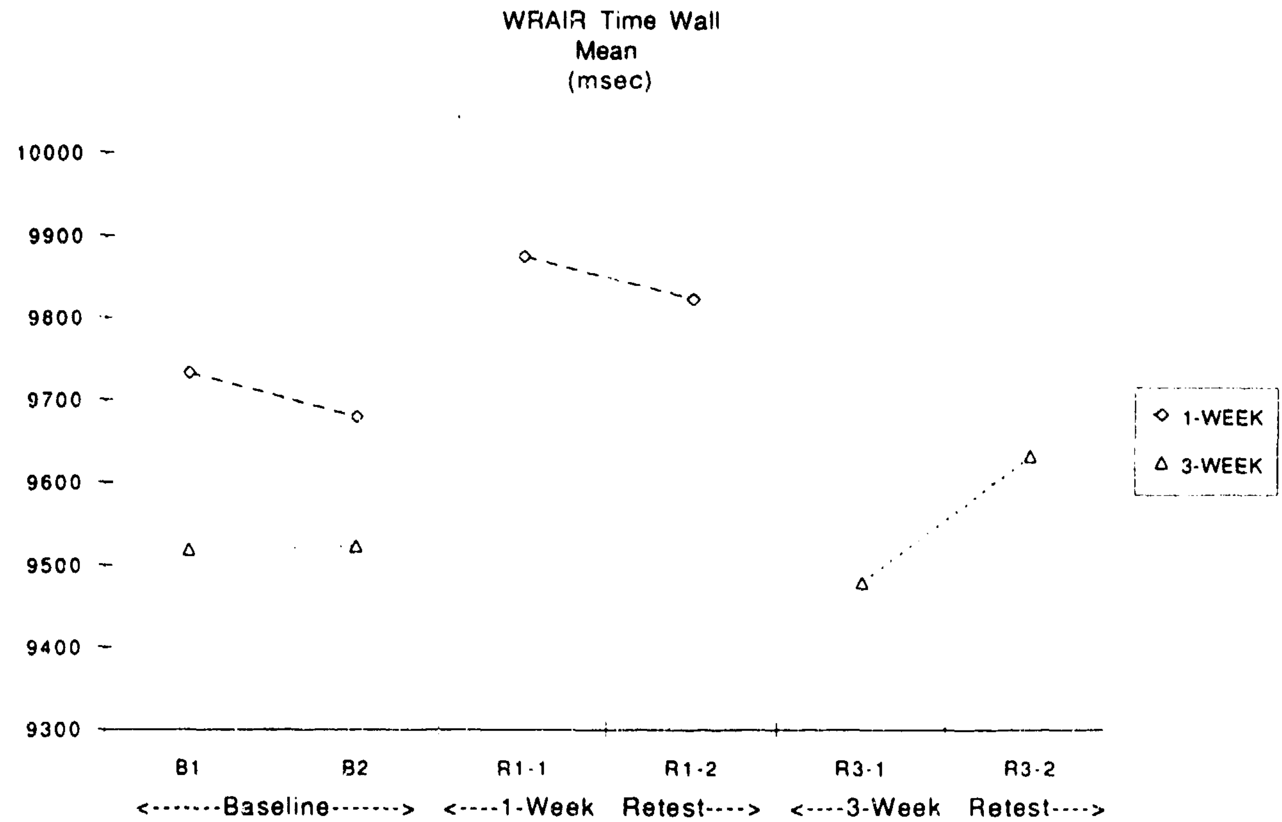


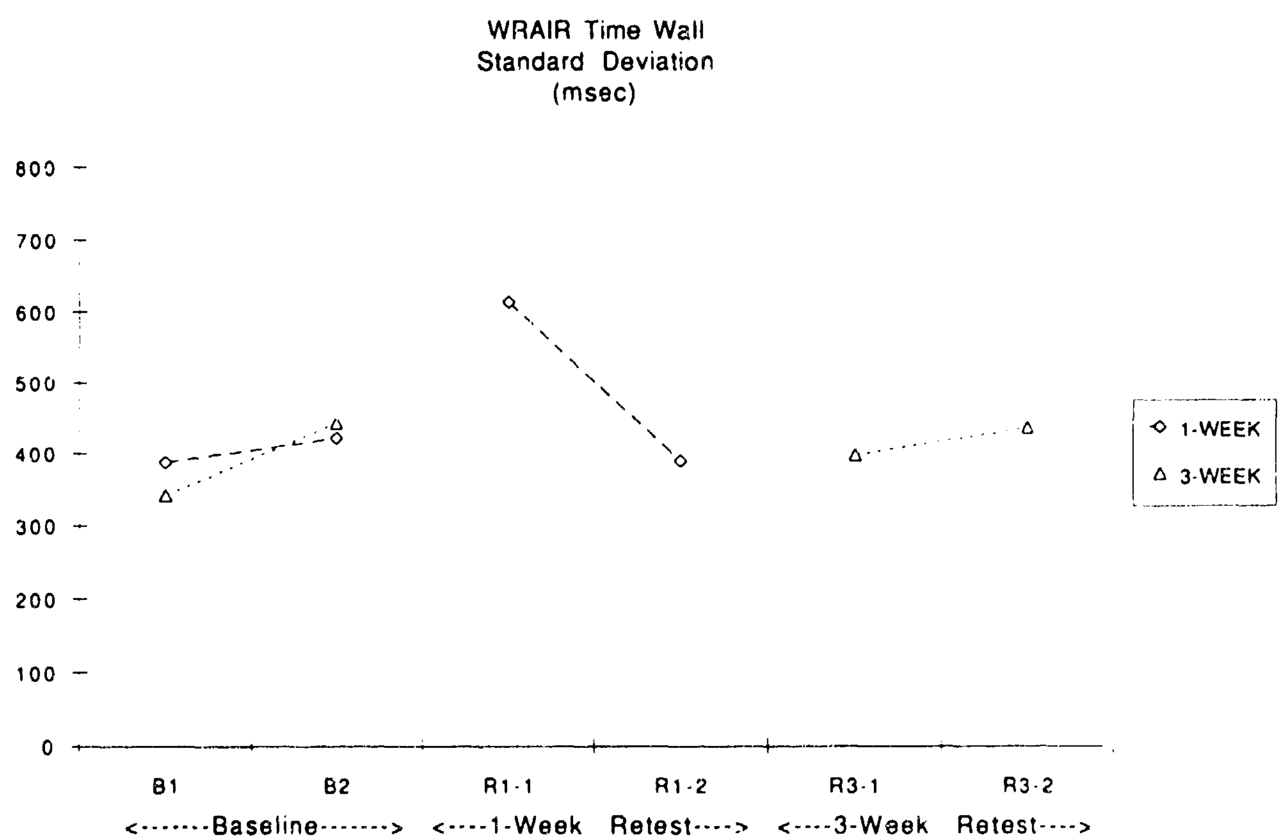
WRAIR Manikin
Mean Response Time
(msec)



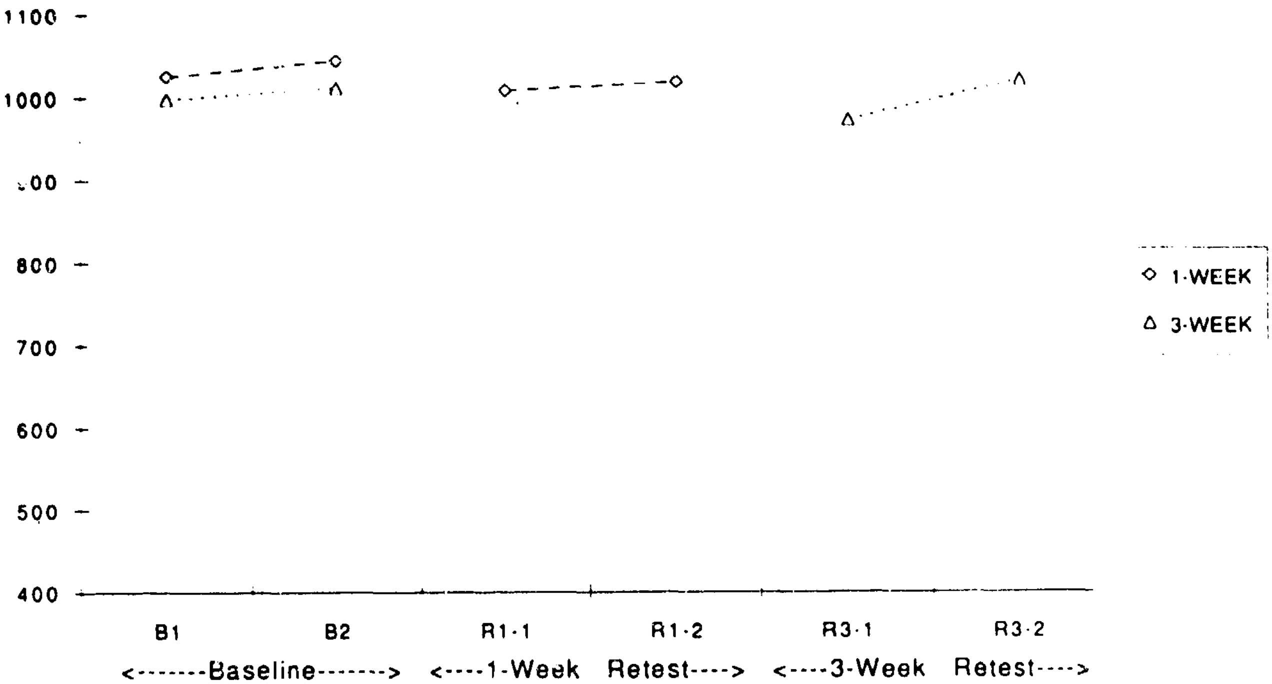
WRAIR Manikin
Percent Correct



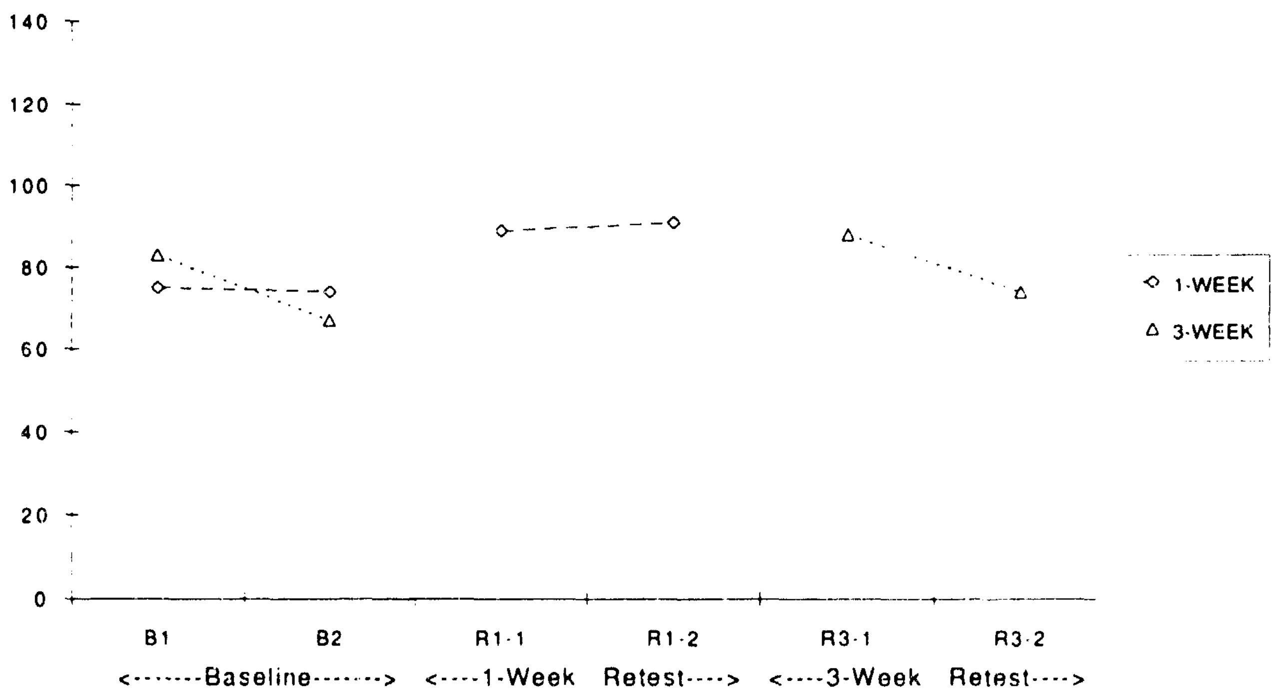

WRAIR Time Wall
Mean
(msec)
10000
9900
9800
9700
9600
9500
9400
9300
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------> <----1-Week Retest----> <----3-Week Retest---->
1-WEEK
3-WEEK


WRAIR Time Wall
Standard Deviation
(msec)
800
700
600
500
400
300
200
100
0
B1
B2
R1-1
R1-2
R3-1
R3-2
<-------Baseline-------> <----1-Week Retest----> <----3-Week Retest---->
1-WEEK
3-WEEK

WRAIR Interval Production
Mean
(msec)

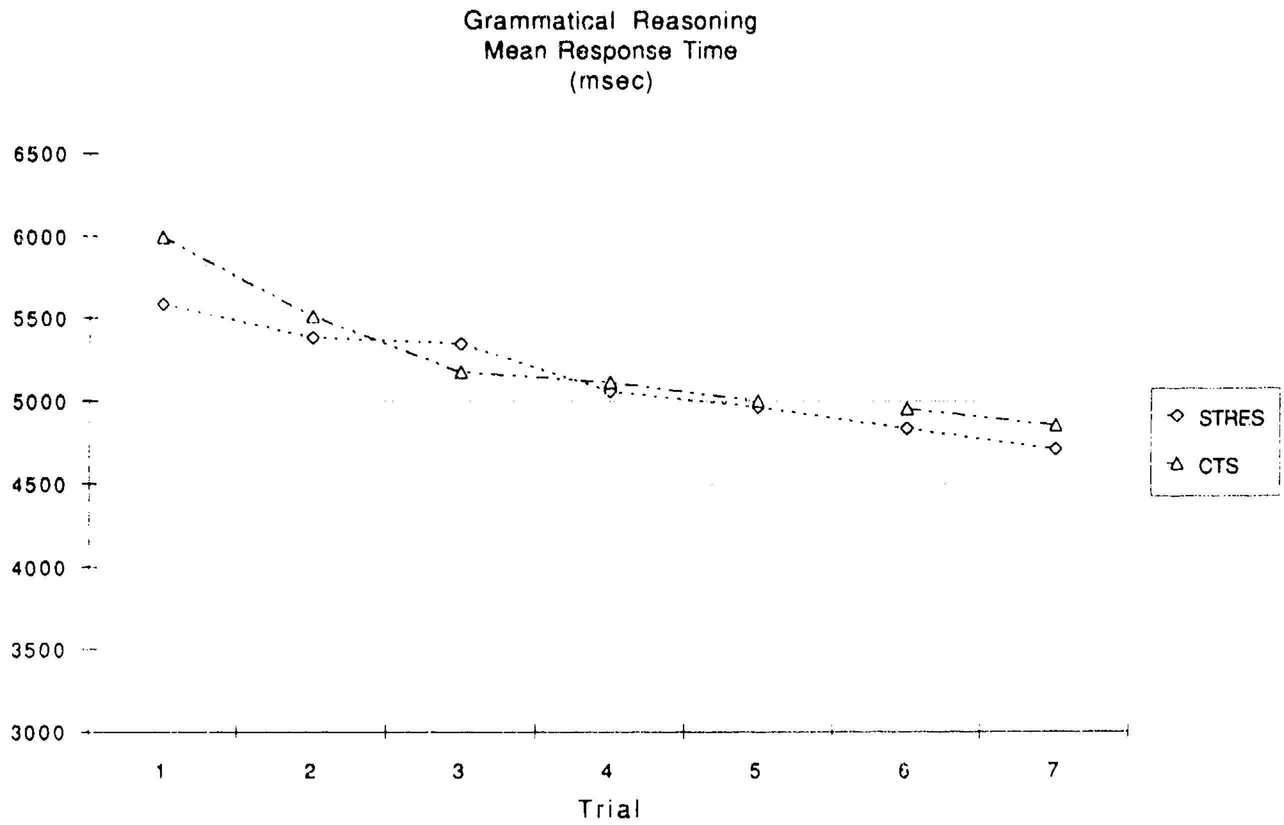


WRAIR Interval Production
Standard Deviation
(msec)

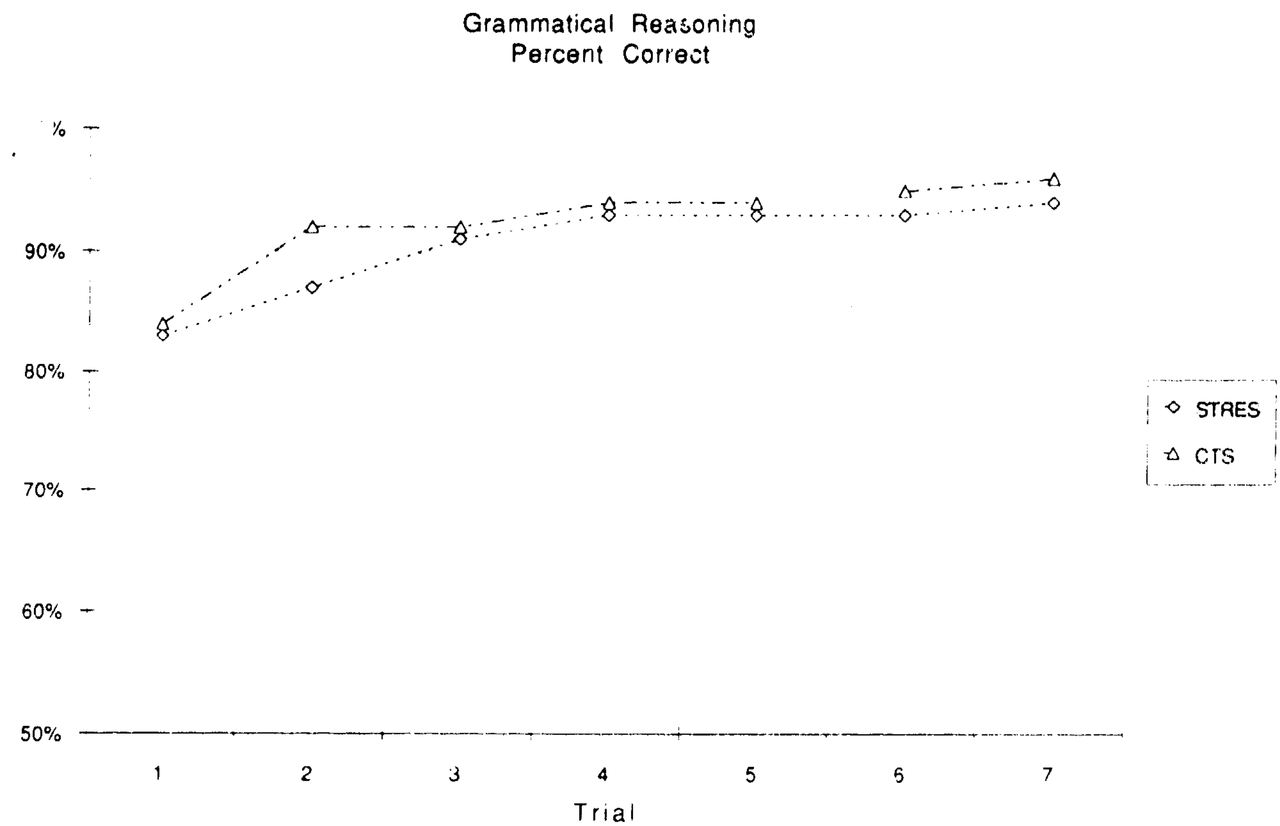


APPENDIX G

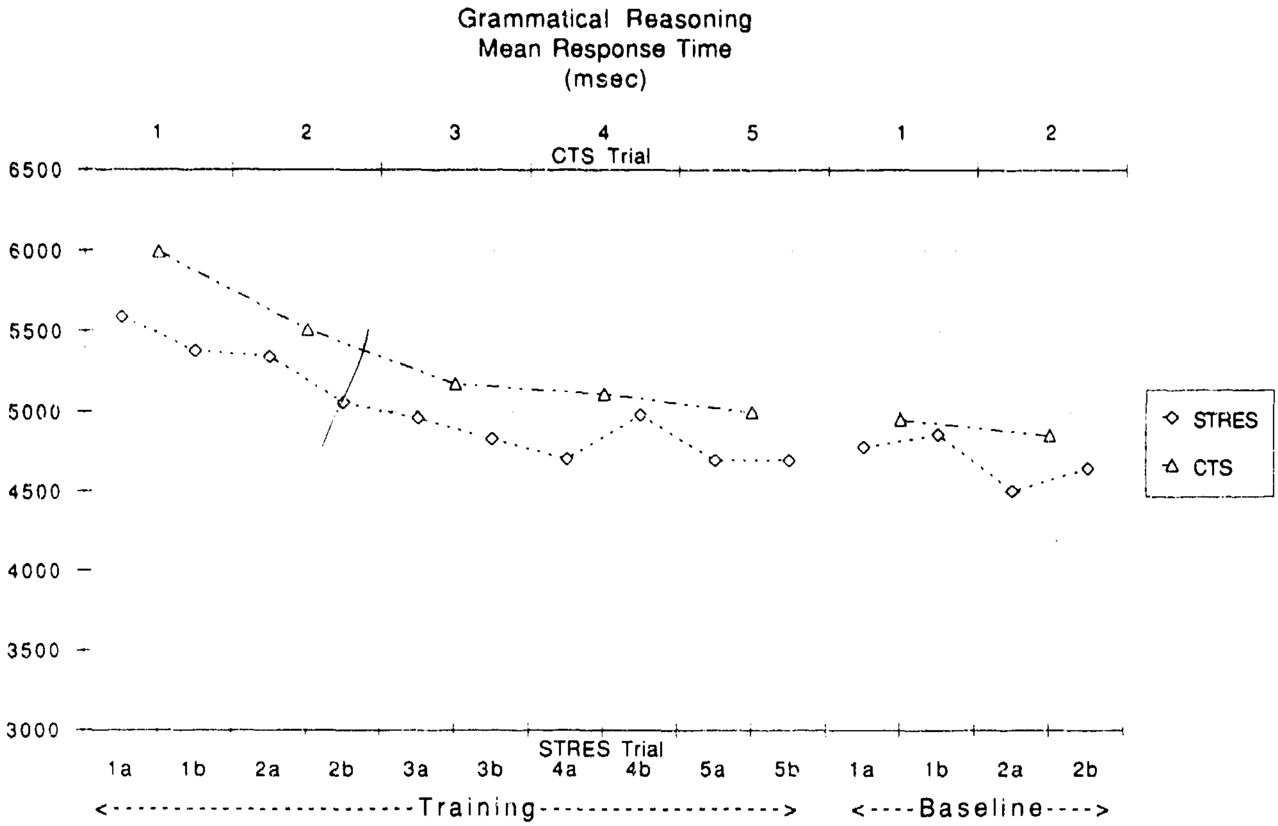
STRES VS. CTS COMPARISON

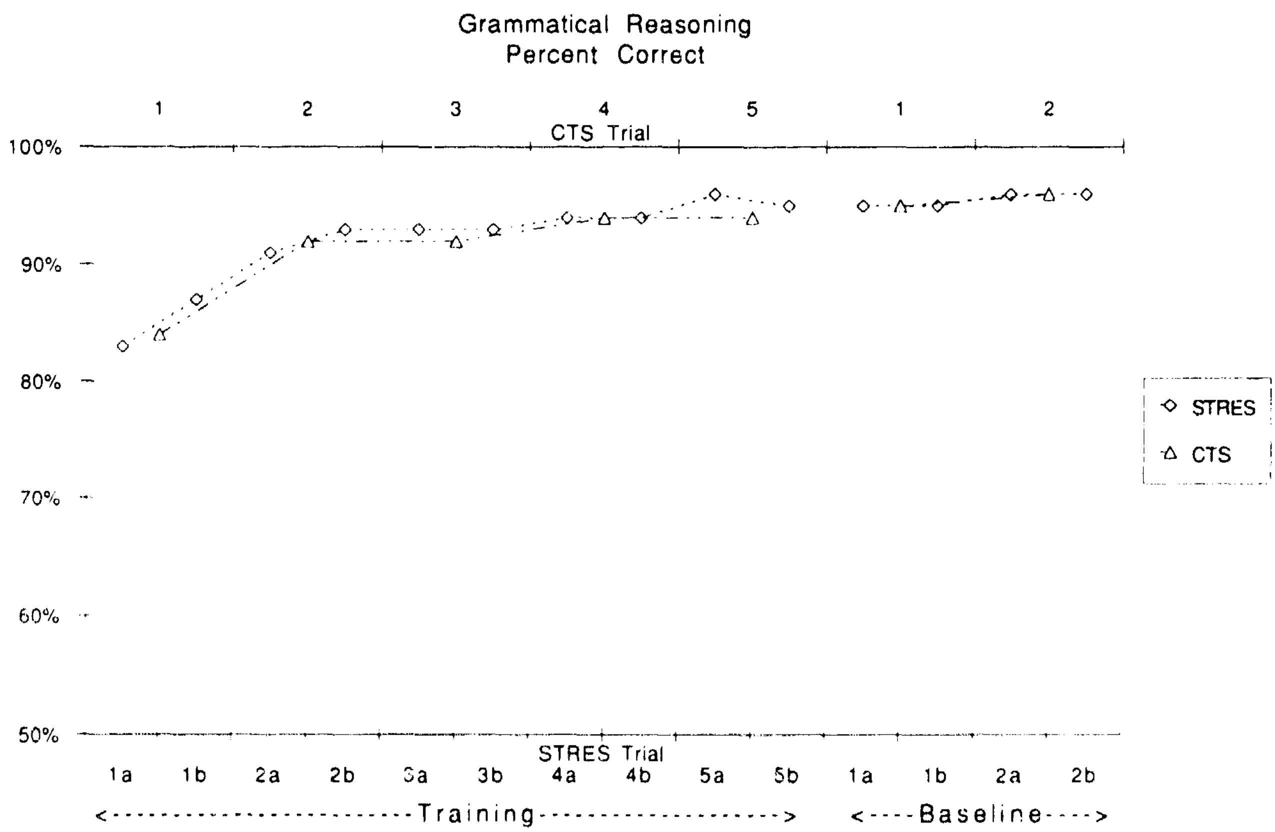


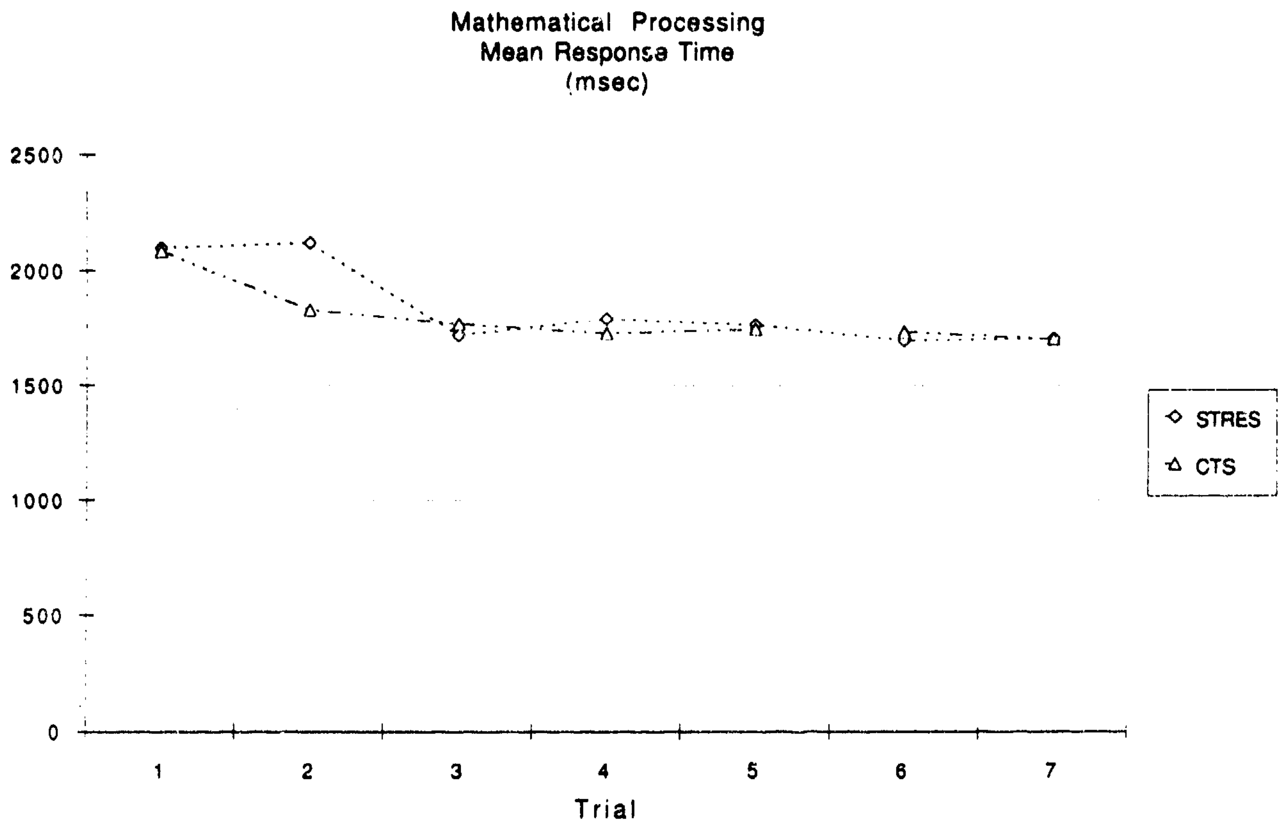
Grammatical Reasoning
Mean Response Time
(msec)
6500
6000
5500
5000
4500
4000
3500
3000
1
2
3
4
5
6
7
Trial
STRES
CTS



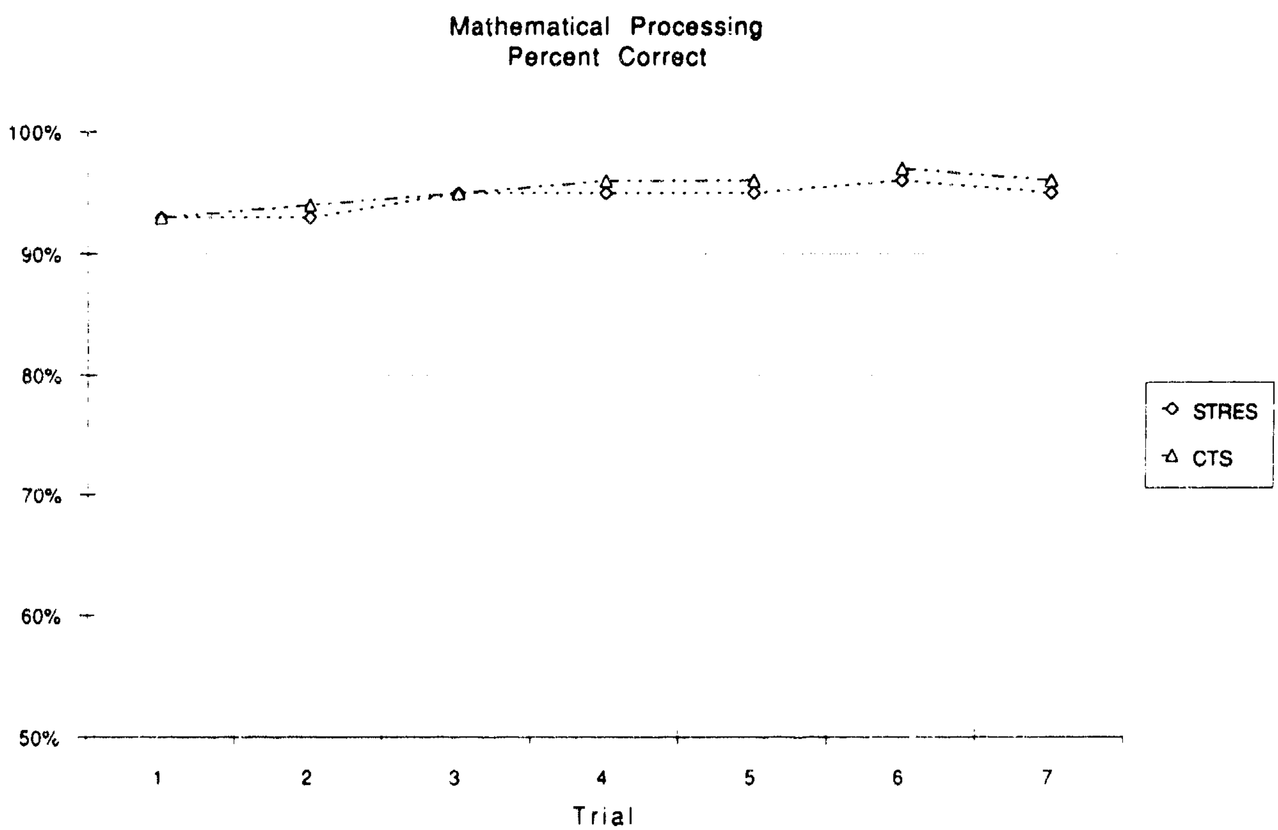
Grammatical Reasoning
Percent Correct
90%
80%
70%
60%
50%
1
2
3
4
5
6
7
Trial
STRES
CTS


Grammatical Reasoning
Mean Response Time
(msec)
1 2 3 4 5 1 2
CTS Trial
6500
6000
5500
5000
4500
4000
3500
3000
STRES Trial
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<----------------------Training-------------------> <----Baseline---->
STRES
CTS

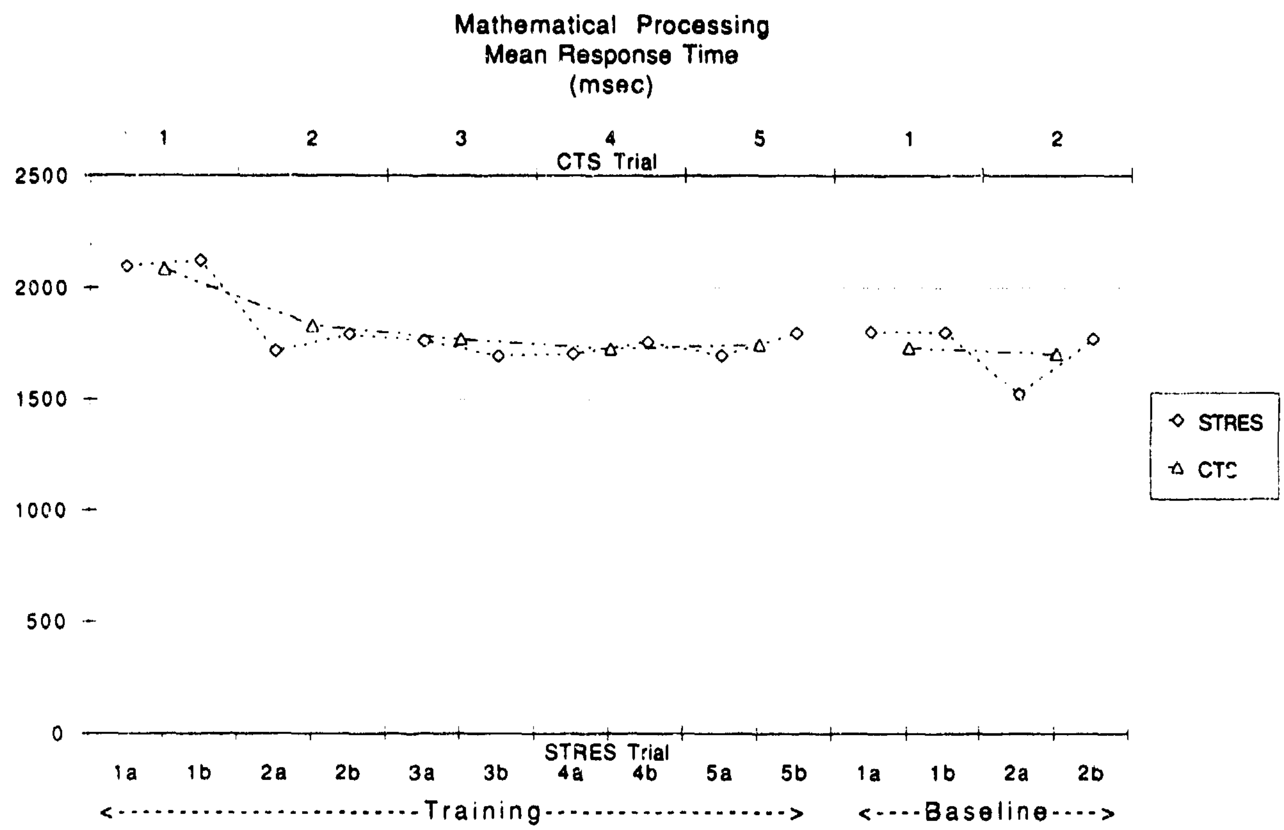

Grammatical Reasoning
Percent Correct
1 2 3 4 5 1 2
CTS Trial
100%
90%
80%
70%
60%
50%
STRES Trial
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<----------------------Training-------------------> <----Baseline---->
STRES
CTS



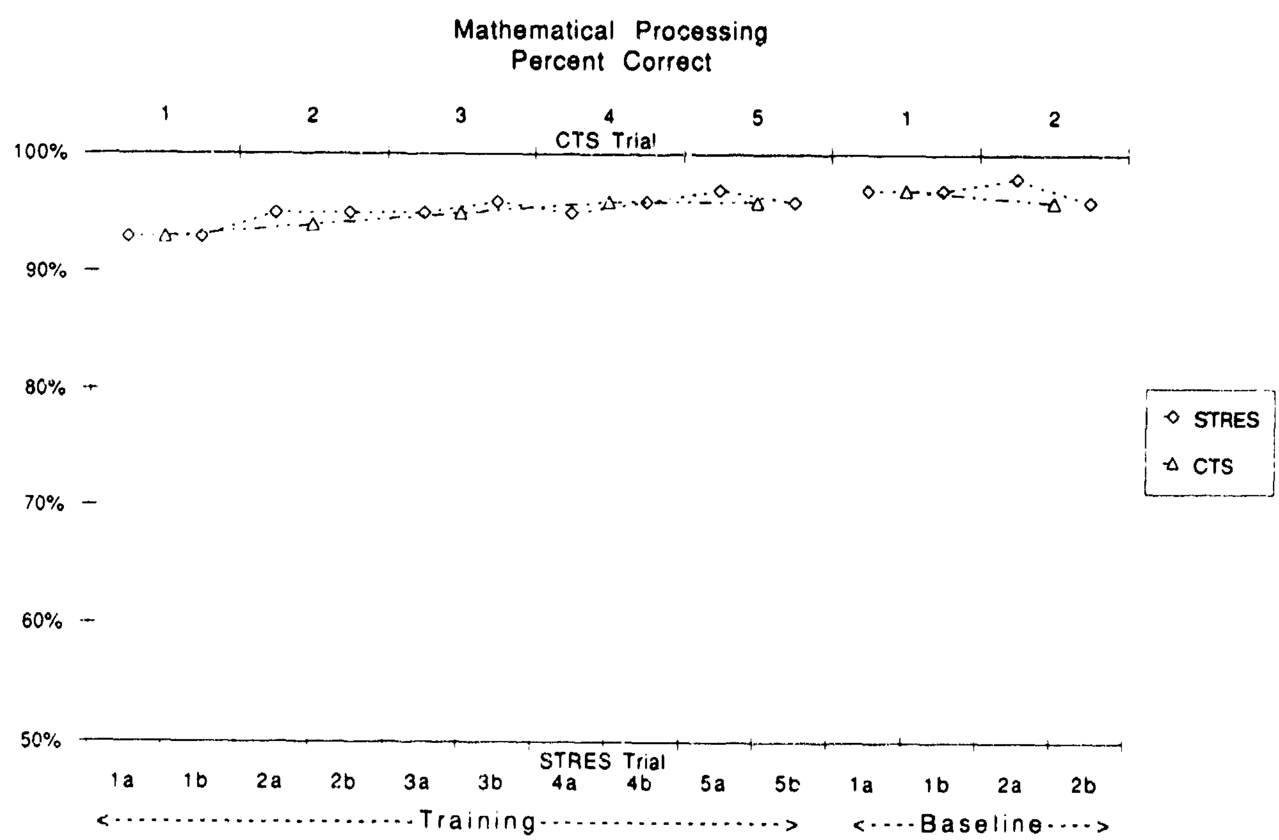
Mathematical Processing
Mean Response Time
(msec)
2500
2000
1500
1000
500
0
1
2
3
4
5
6
7
Trial
STRES
CTS



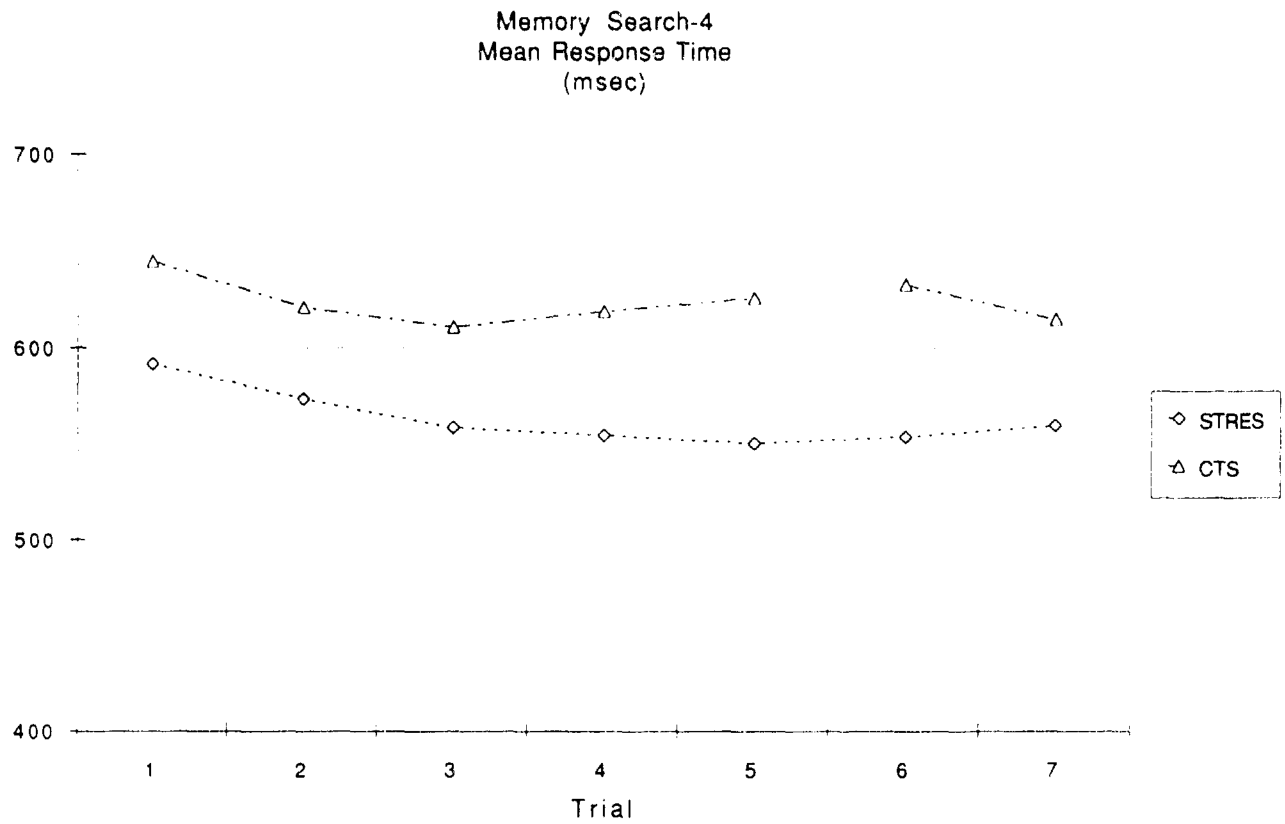
Mathematical Processing
Percent Correct
100%
90%
80%
70%
60%
50%
1
2
3
4
5
6
7
Trial
STRES
CTS



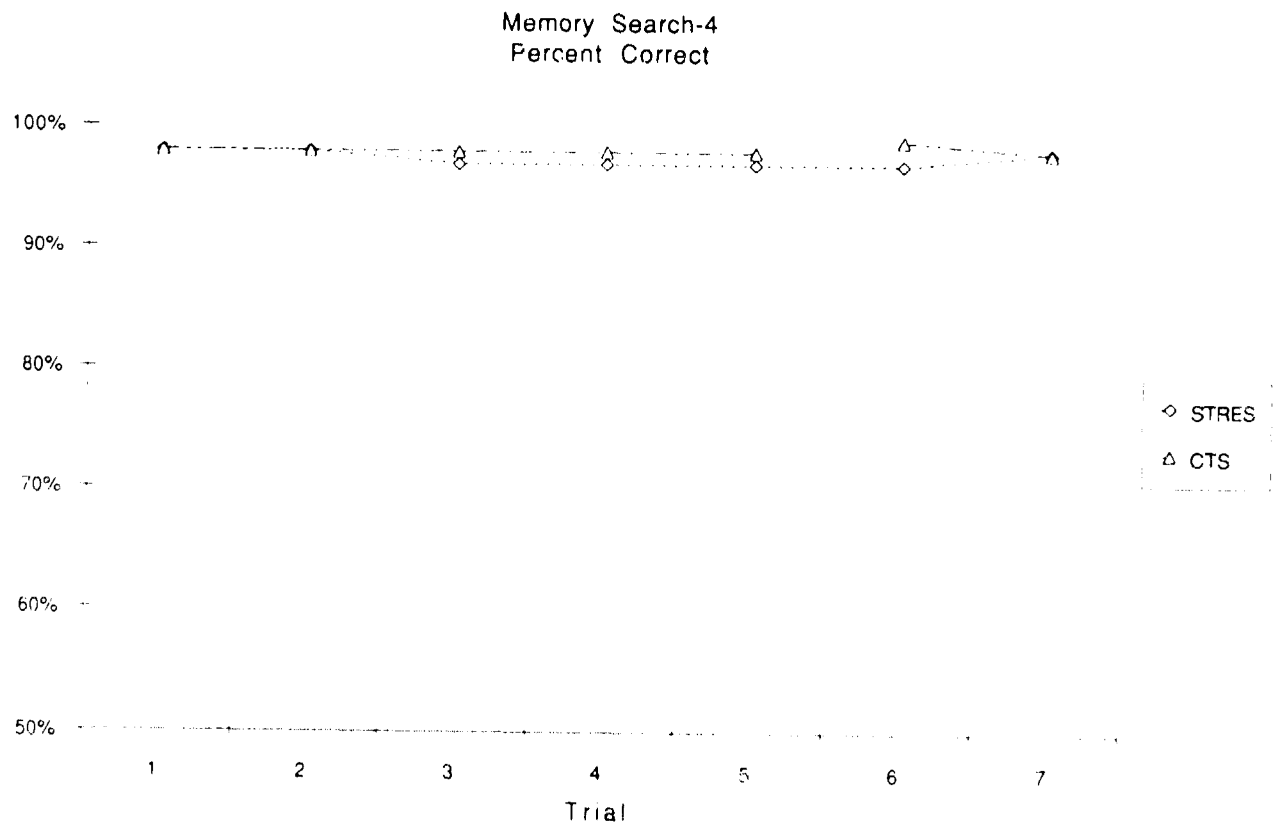
Mathematical Processing
Mean Response Time
(msec)
CTS Trial
1 2 3 4 5 1 2
2500
2000
1500
1000
500
0
STRES Trial
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<------Training------> <----Baseline---->
STRES
CTS



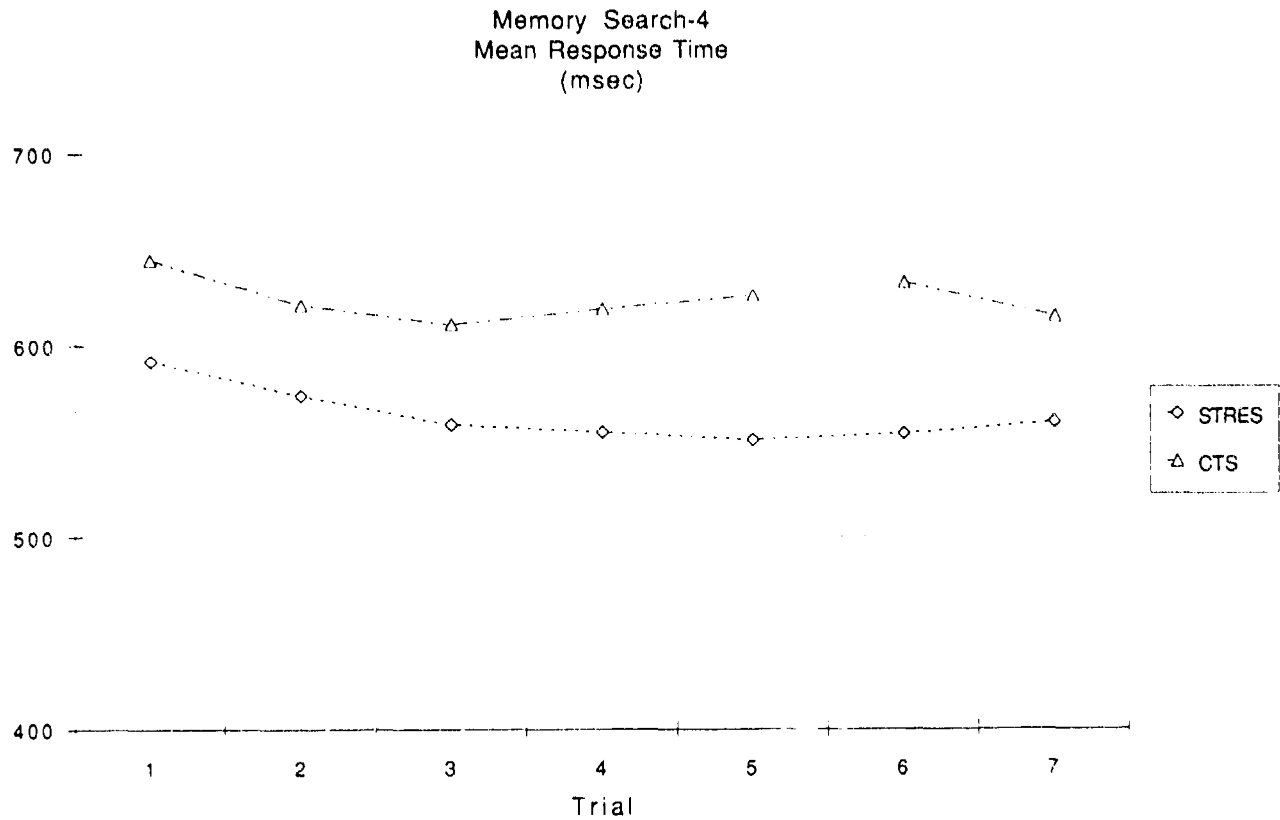
Mathematical Processing
Percent Correct
CTS Trial
1 2 3 4 5 1 2
100%
90%
80%
70%
60%
50%
STRES Trial
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<------Training------> <----Baseline---->
STRES
CTS



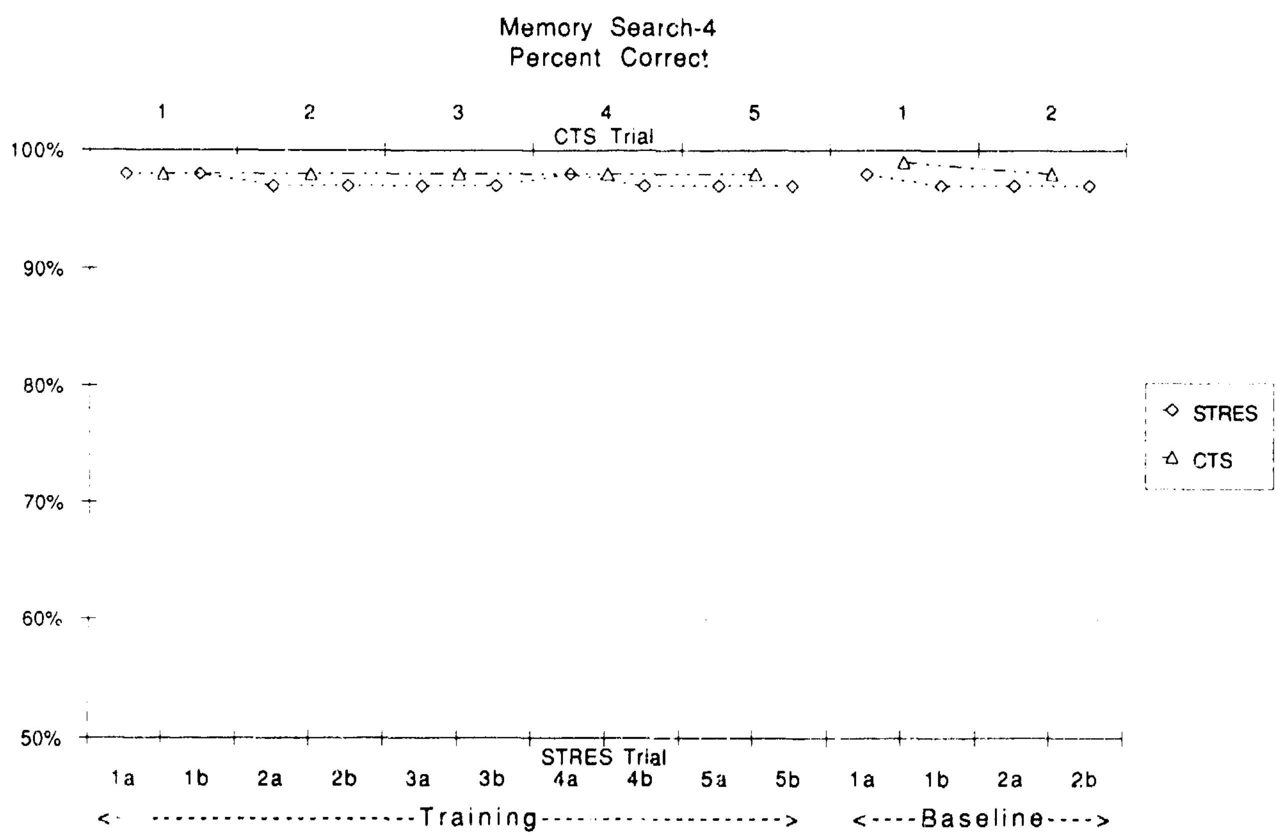
Memory Search-4
Mean Response Time
(msec)
700
600
500
400
1
2
3
4
5
6
7
Trial
STRES
CTS



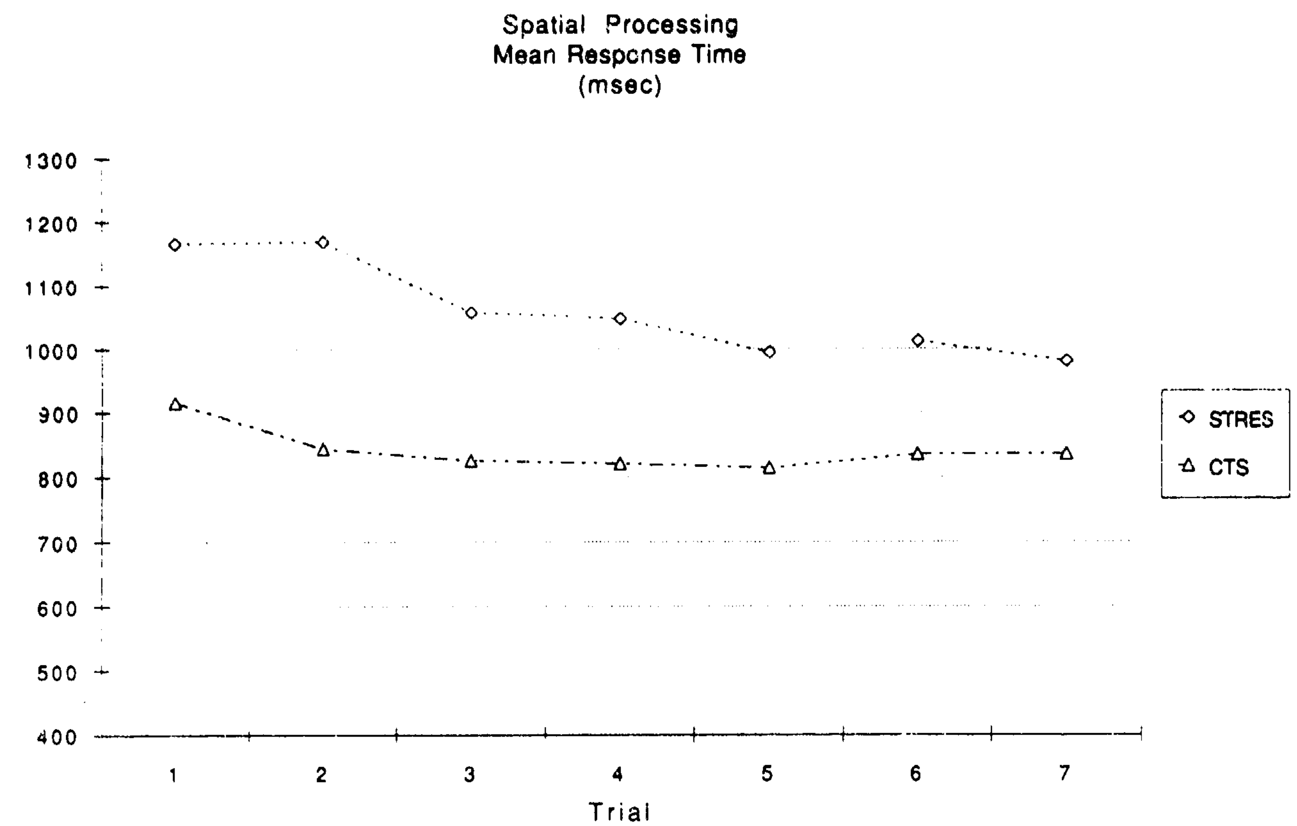
Memory Search-4
Percent Correct
100%
90%
80%
70%
60%
50%
1
2
3
4
5
6
7
Trial
STRES
CTS



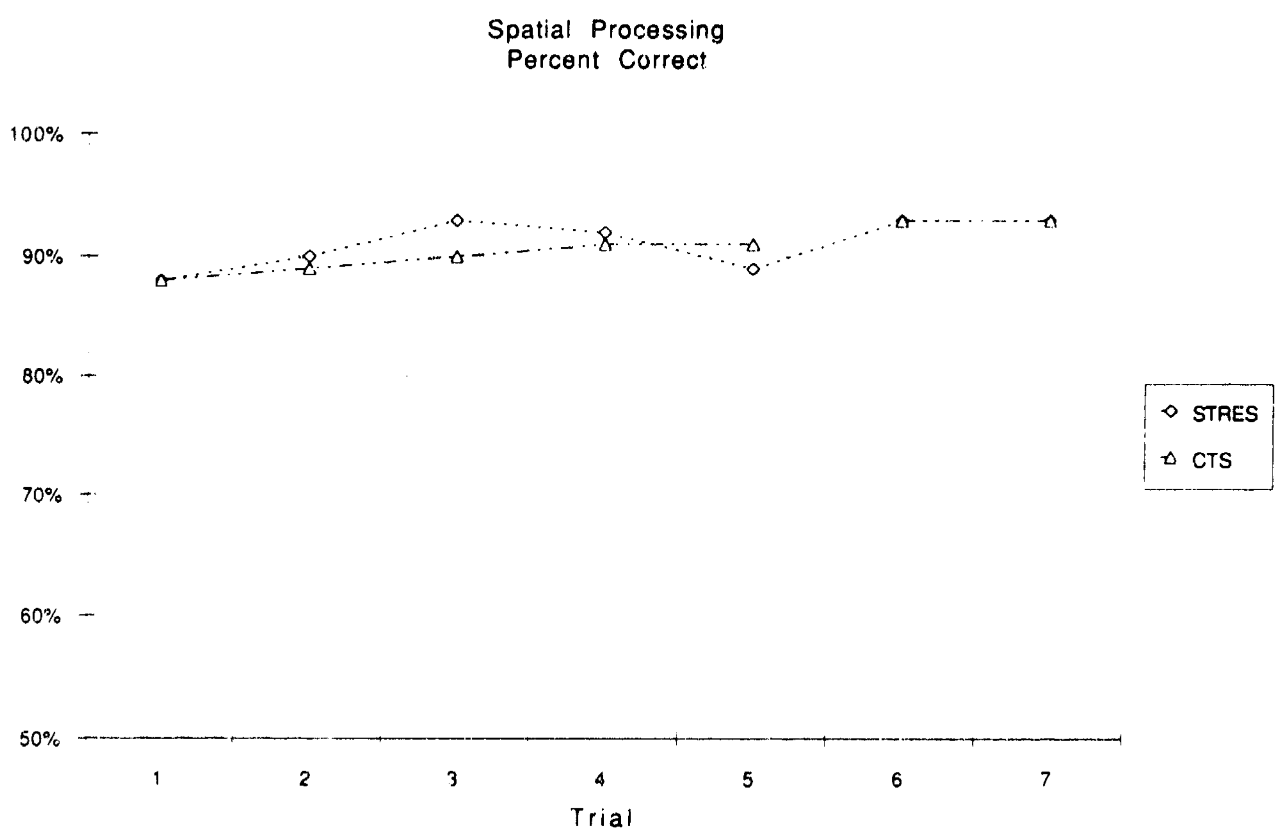
Memory Search-4
Mean Response Time
(msec)
700
600
500
400
1
2
3
4
5
6
7
Trial
STRES
CTS



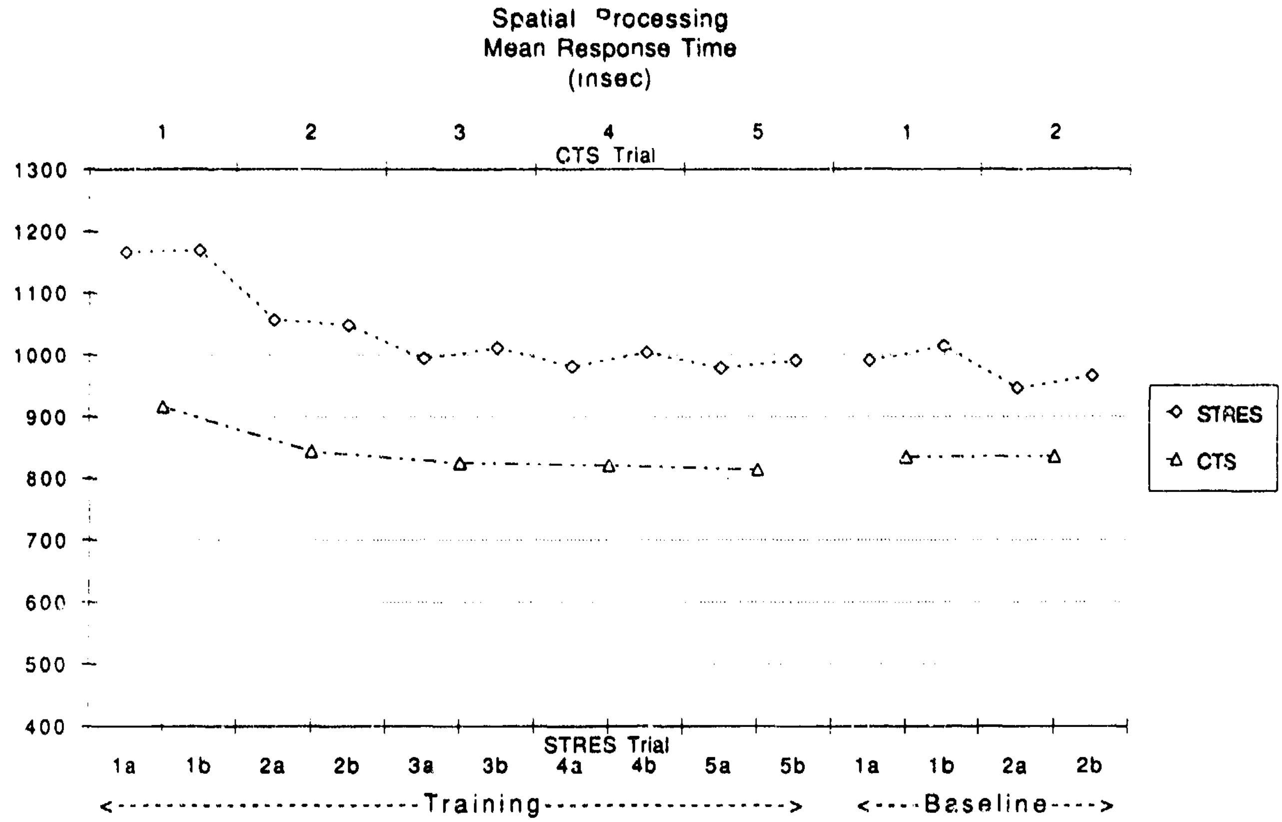
Memory Search-4
Percent Correct
1
2
3
4
5
1
2
CTS Trial
100%
90%
80%
70%
60%
50%
STRES Trial
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<--------------------Training-------------------->
<----Baseline---->
STRES
CTS



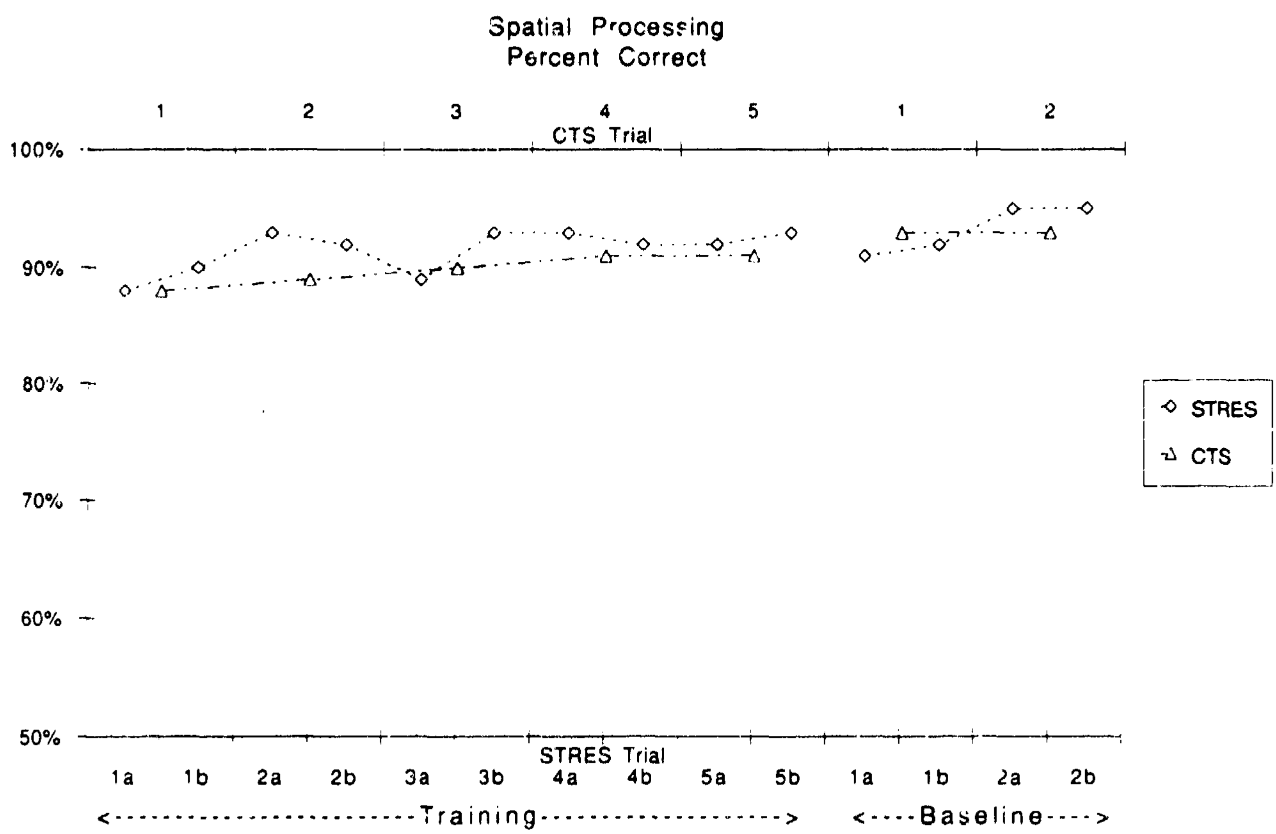
Spatial Processing
Mean Response Time
(msec)
1300
1200
1100
1000
900
800
700
600
500
400
1
2
3
4
5
6
7
Trial
STRES
CTS



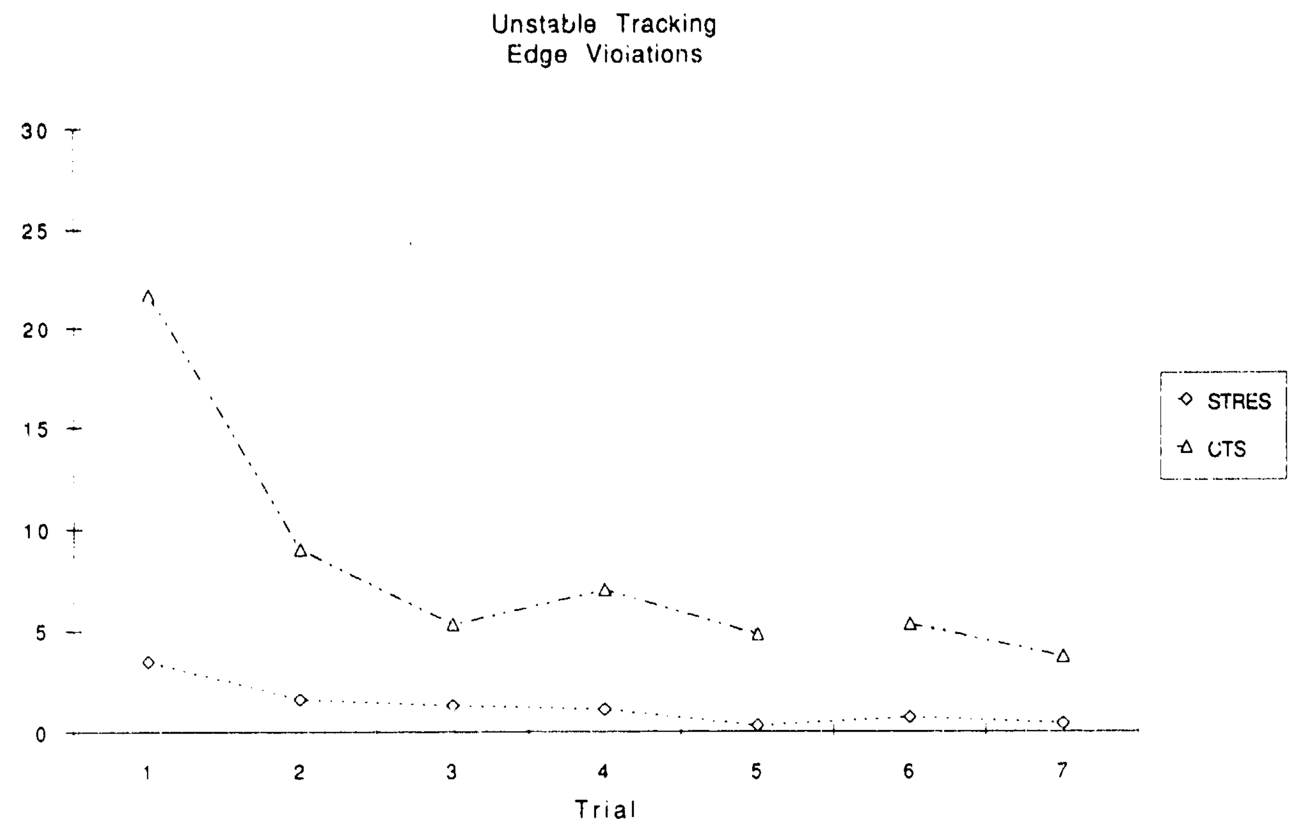
Spatial Processing
Percent Correct
100%
90%
80%
70%
60%
50%
1
2
3
4
5
6
7
Trial
STRES
CTS

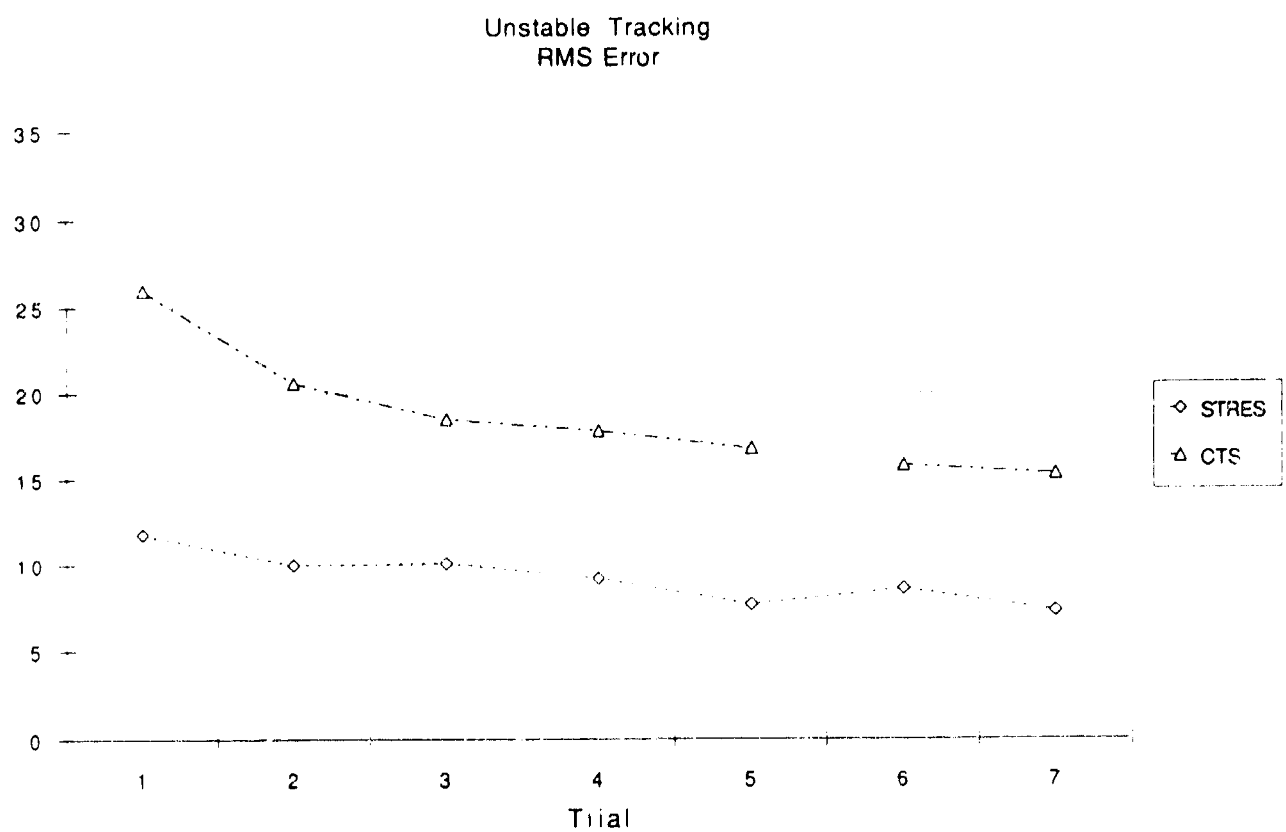


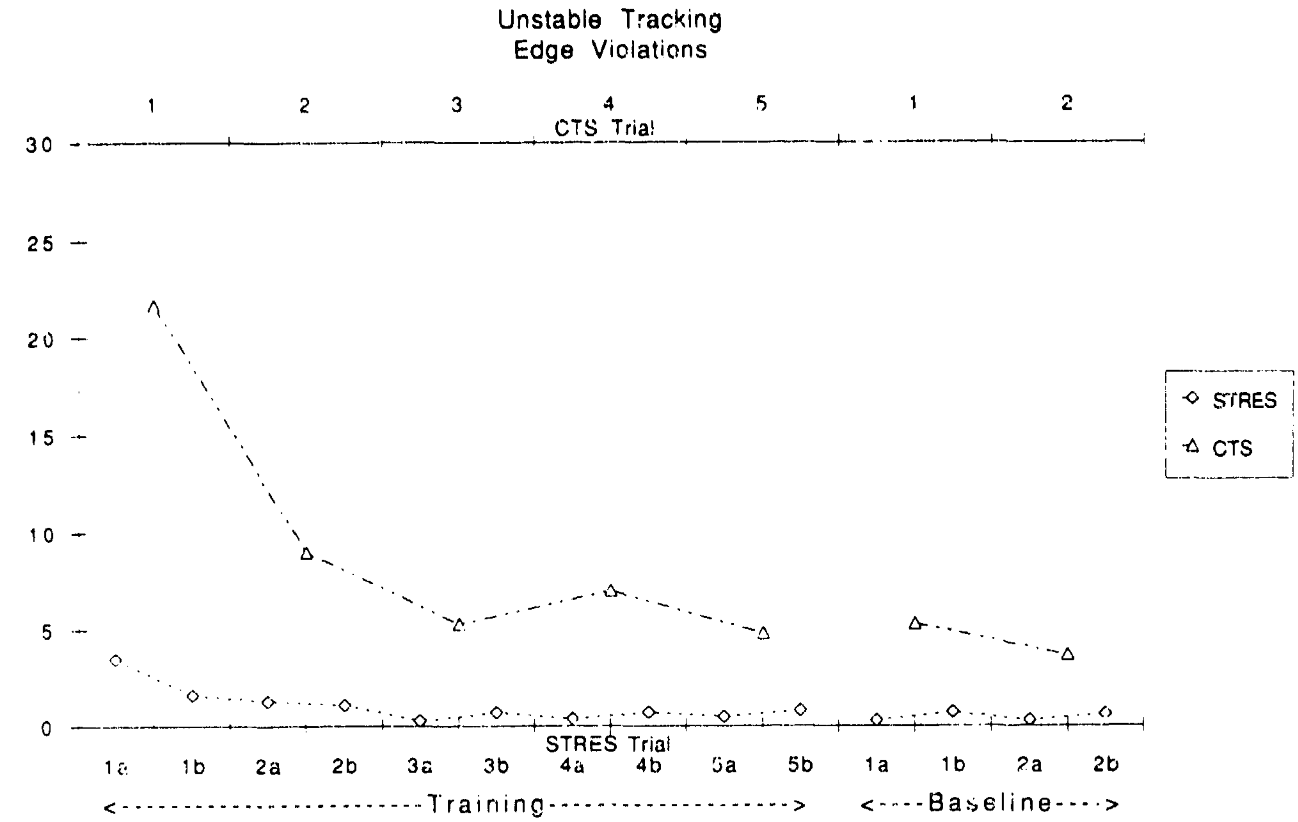
Spatial Processing
Mean Response Time
(msec)
1 2 3 4 5 1 2
CTS Trial
1300
1200
1100
1000
900
800
700
600
500
400
STRES Trial
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<----------Training----------> <----Baseline---->
STRES
CTS

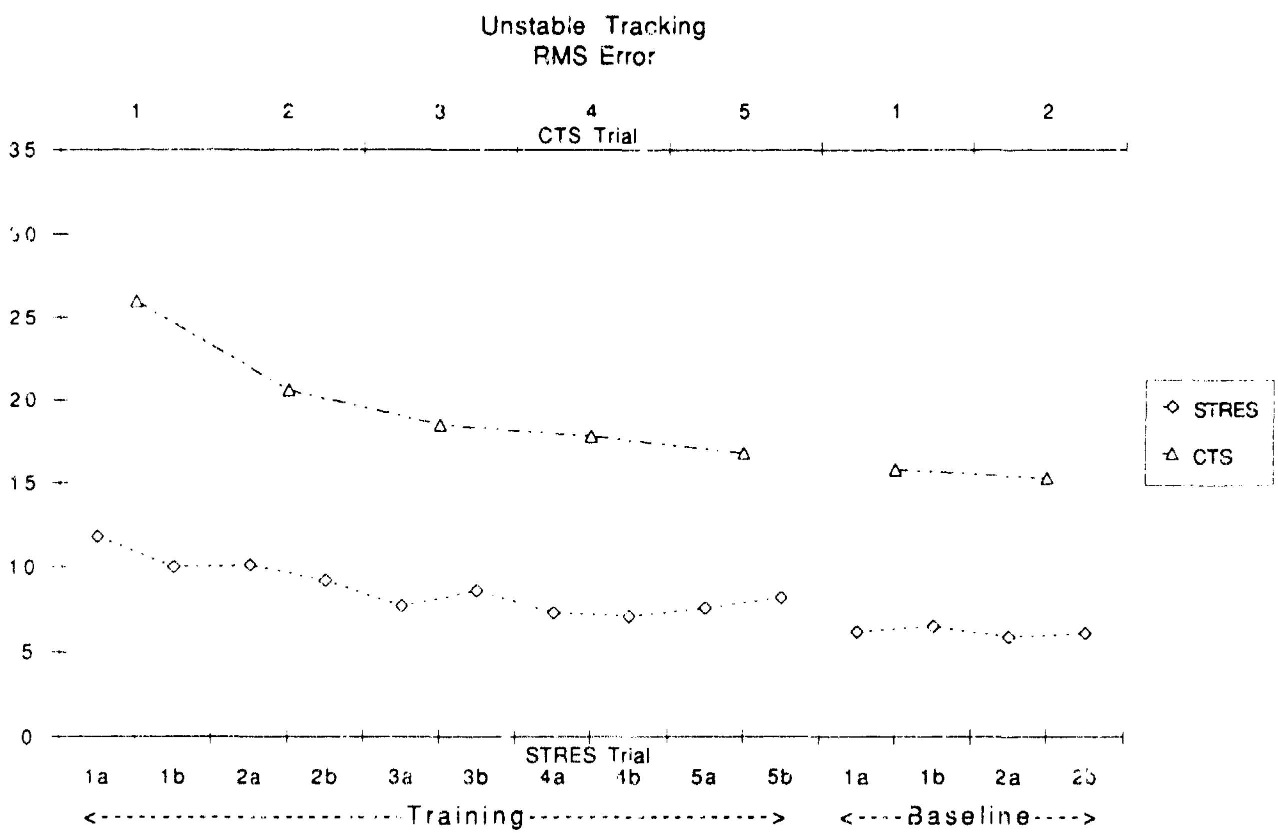


Spatial Processing
Percent Correct
1 2 3 4 5 1 2
CTS Trial
100%
90%
80%
70%
60%
50%
STRES Trial
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<----------Training----------> <----Baseline---->
STRES
CTS


Unstable Tracking
Edge Violations
30
25
20
15
10
5
0
1
2
3
4
5
6
7
Trial
STRES
CTS


Unstable Tracking
RMS Error
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
Trial
STRES
CTS

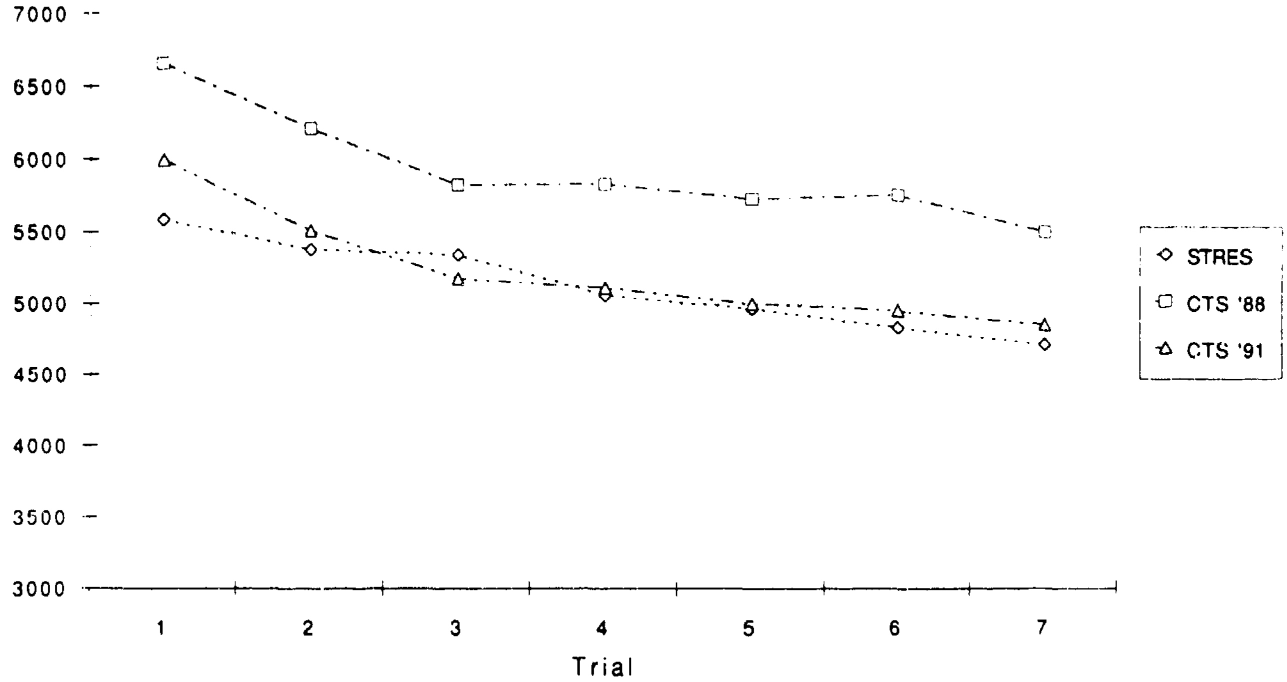

Unstable Tracking
Edge Violations
CTS Trial
1 2 3 4 5 1 2
30
25
20
15
10
5
0
STRES
CTS
STRES Trial
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<----------Training---------->
<----Baseline---->


Unstable Tracking
RMS Error
CTS Trial
1 2 3 4 5 1 2
35
30
25
20
15
10
5
0
STRES
CTS
STRES Trial
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<----------Training---------->
<----Baseline---->

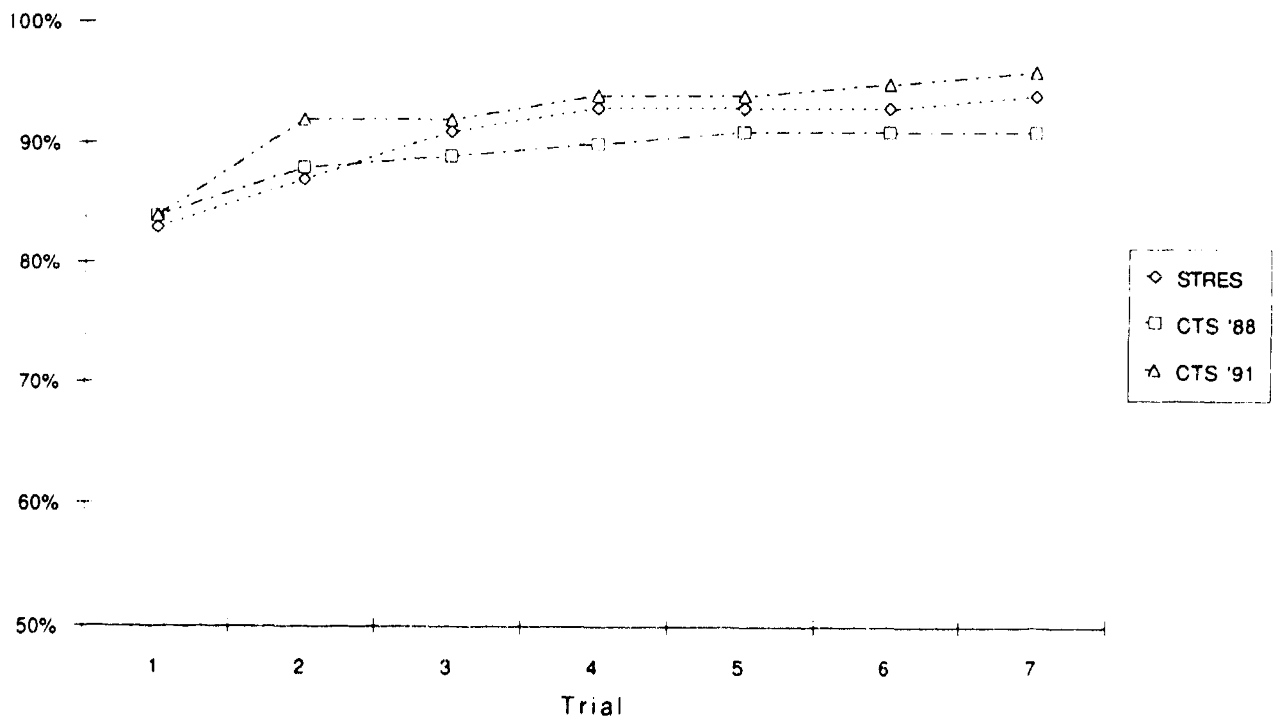
APPENDIX H

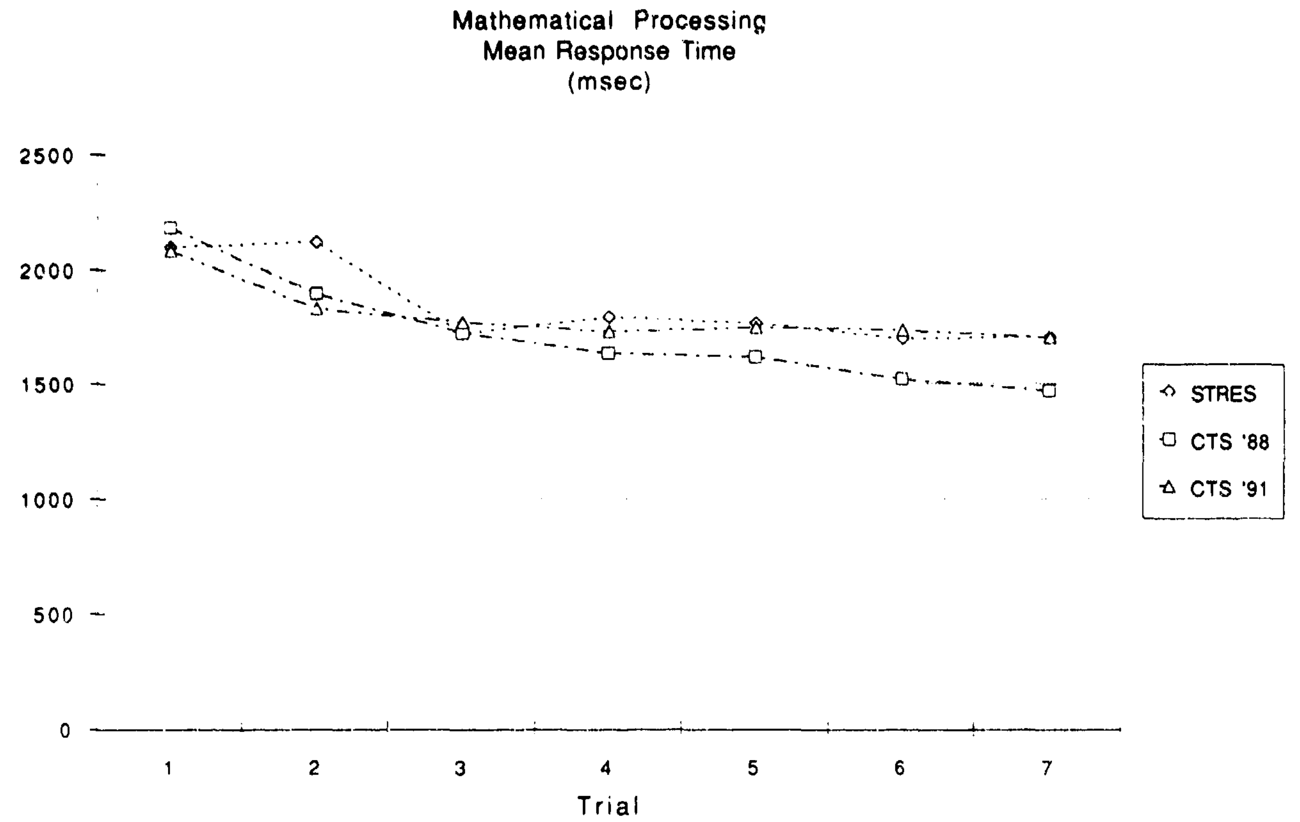
CURRENT AND FORMER CTS DATABASES

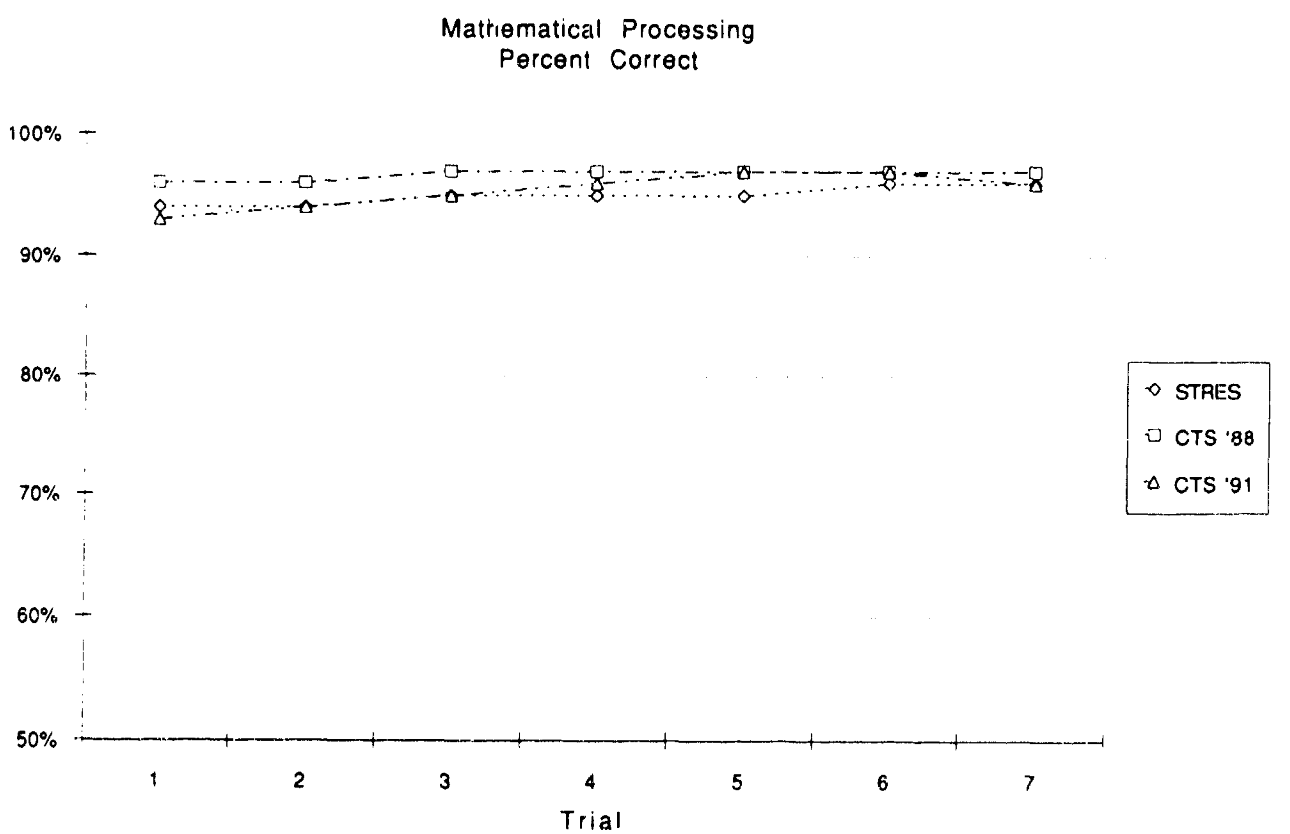
Grammatical Reasoning
Mean Response Time
(msec)

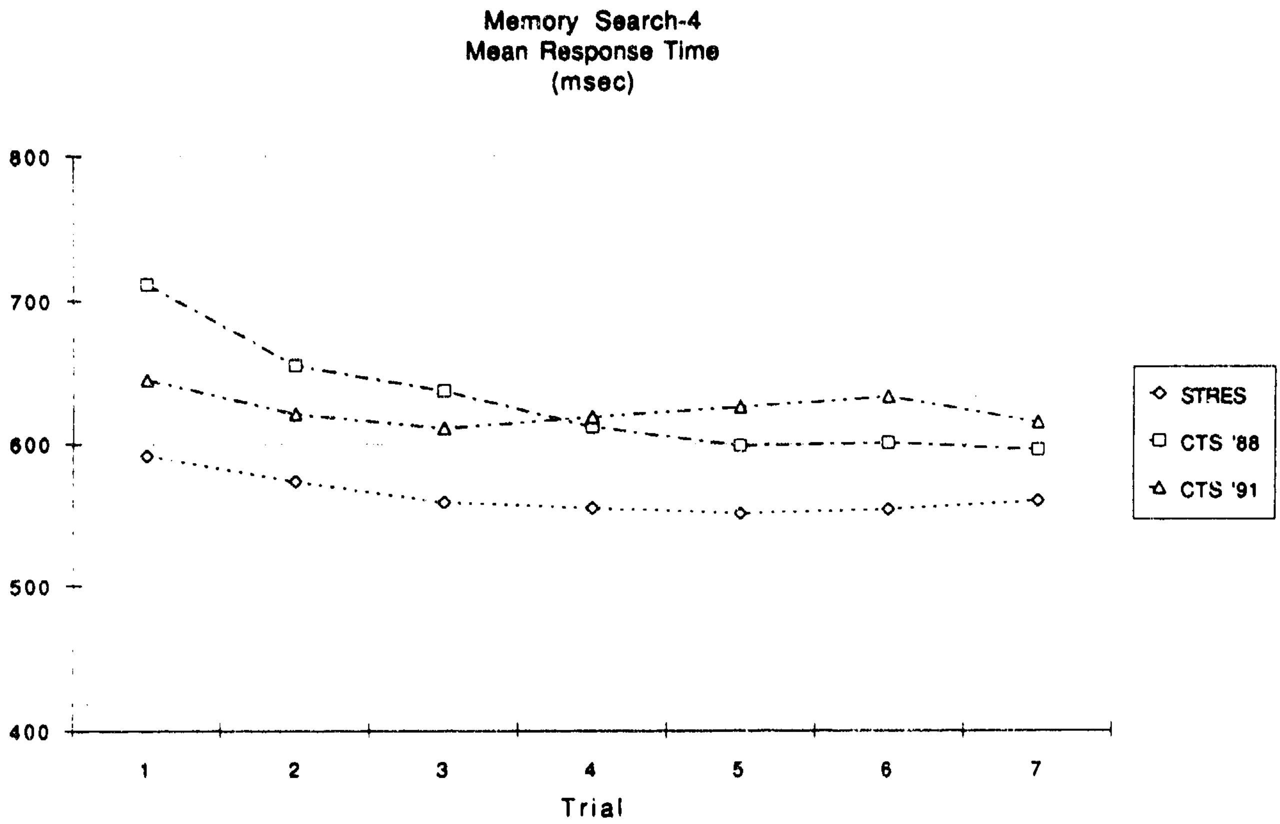


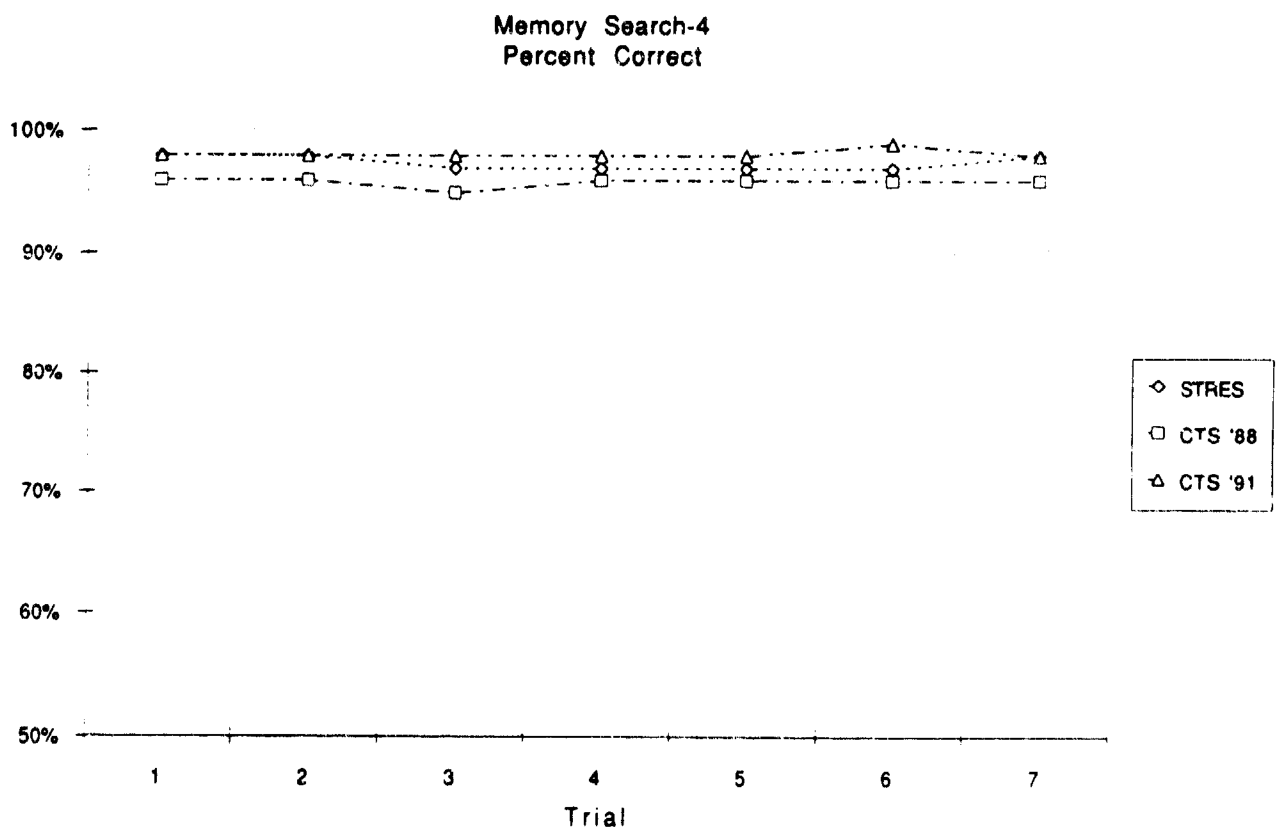
Grammatical Reasoning
Percent Correct

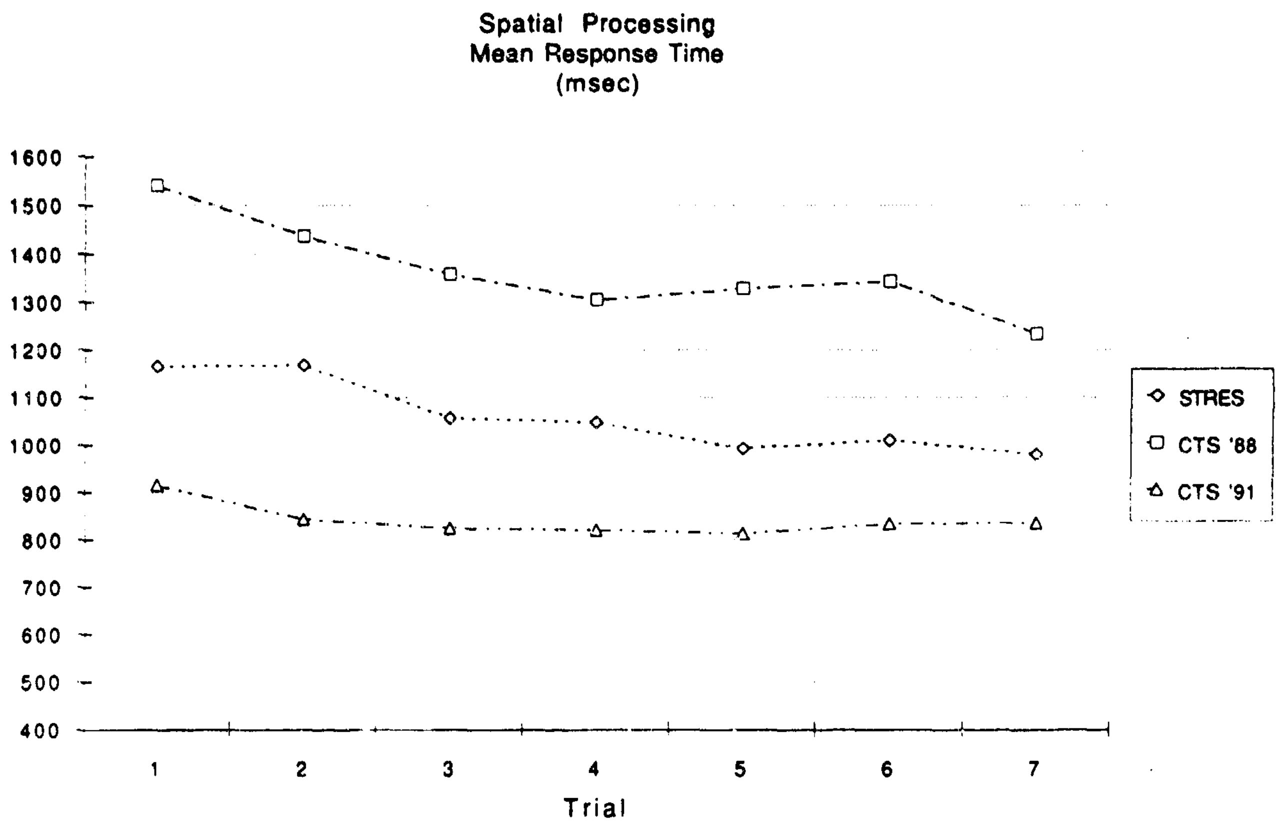


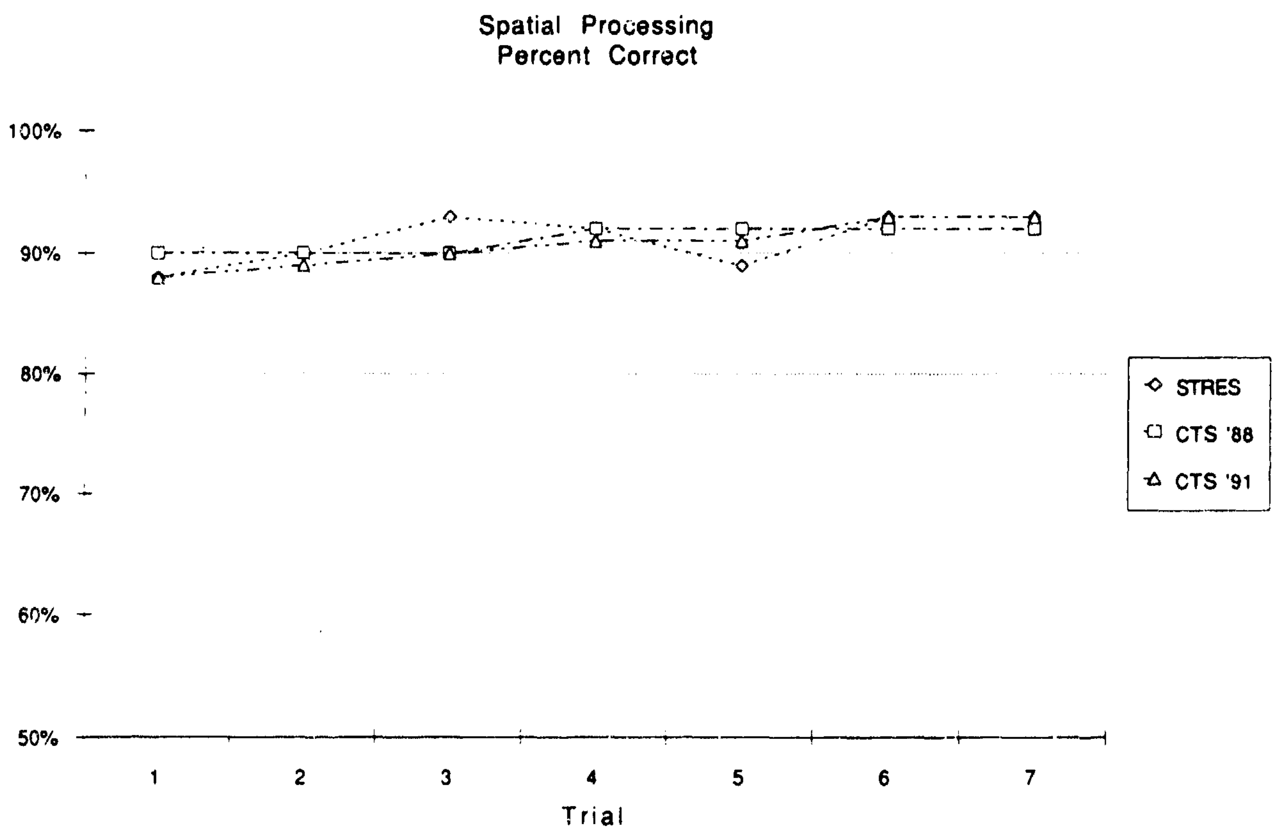

Mathematical Processing
Mean Response Time
(msec)
2500
2000
1500
1000
500
0
1
2
3
4
5
6
7
Trial
STRES
CTS '88
CTS '91

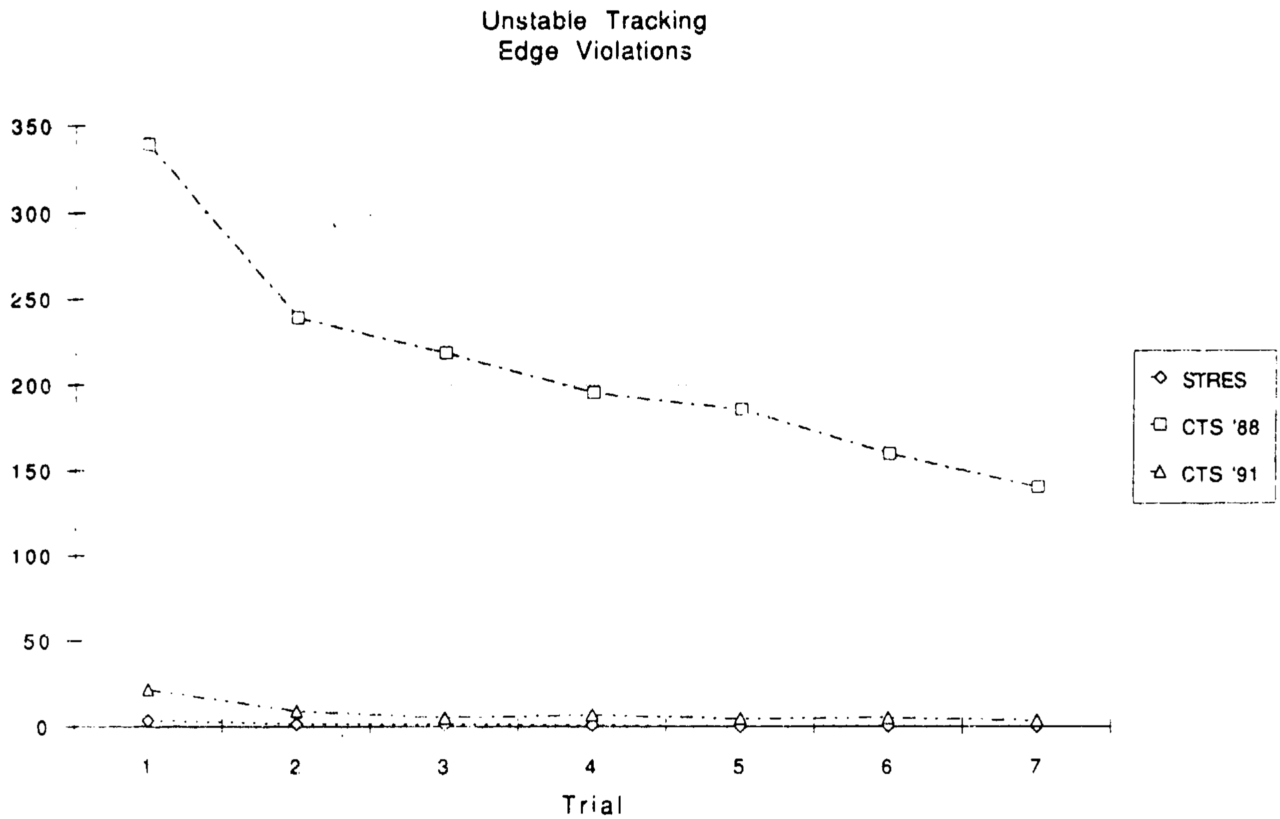

Mathematical Processing
Percent Correct
100%
90%
80%
70%
60%
50%
1
2
3
4
5
6
7
Trial
STRES
CTS '88
CTS '91


Memory Search-4
Mean Response Time
(msec)
800
700
600
500
400
1
2
3
4
5
6
7
Trial
STRES
CTS '88
CTS '91

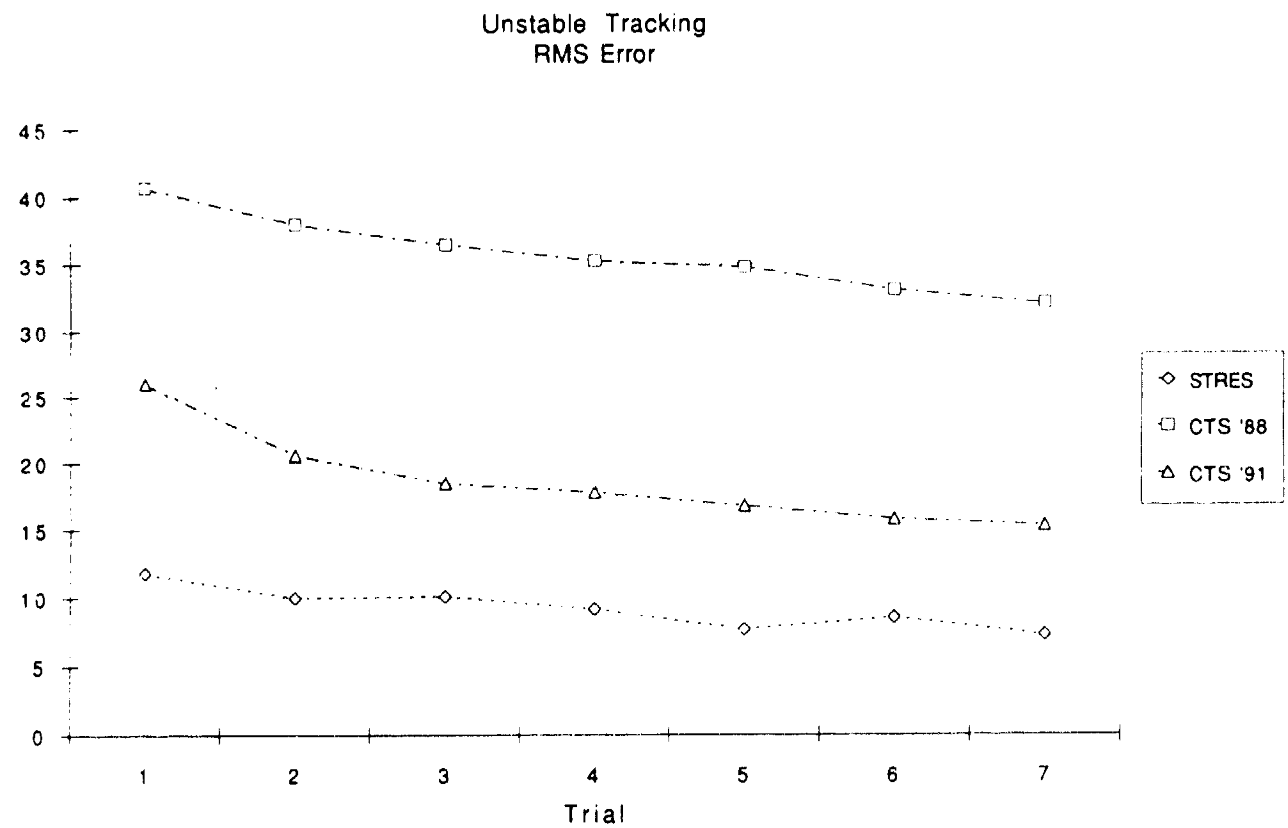

Memory Search-4
Percent Correct
100%
90%
80%
70%
60%
50%
1
2
3
4
5
6
7
Trial
STRES
CTS '88
CTS '91


Spatial Processing
Mean Response Time
(msec)
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
500
400
1
2
3
4
5
6
7
Trial
STRES
CTS '88
CTS '91


Spatial Processing
Percent Correct
100%
90%
80%
70%
60%
50%
1
2
3
4
5
6
7
Trial
STRES
CTS '88
CTS '91



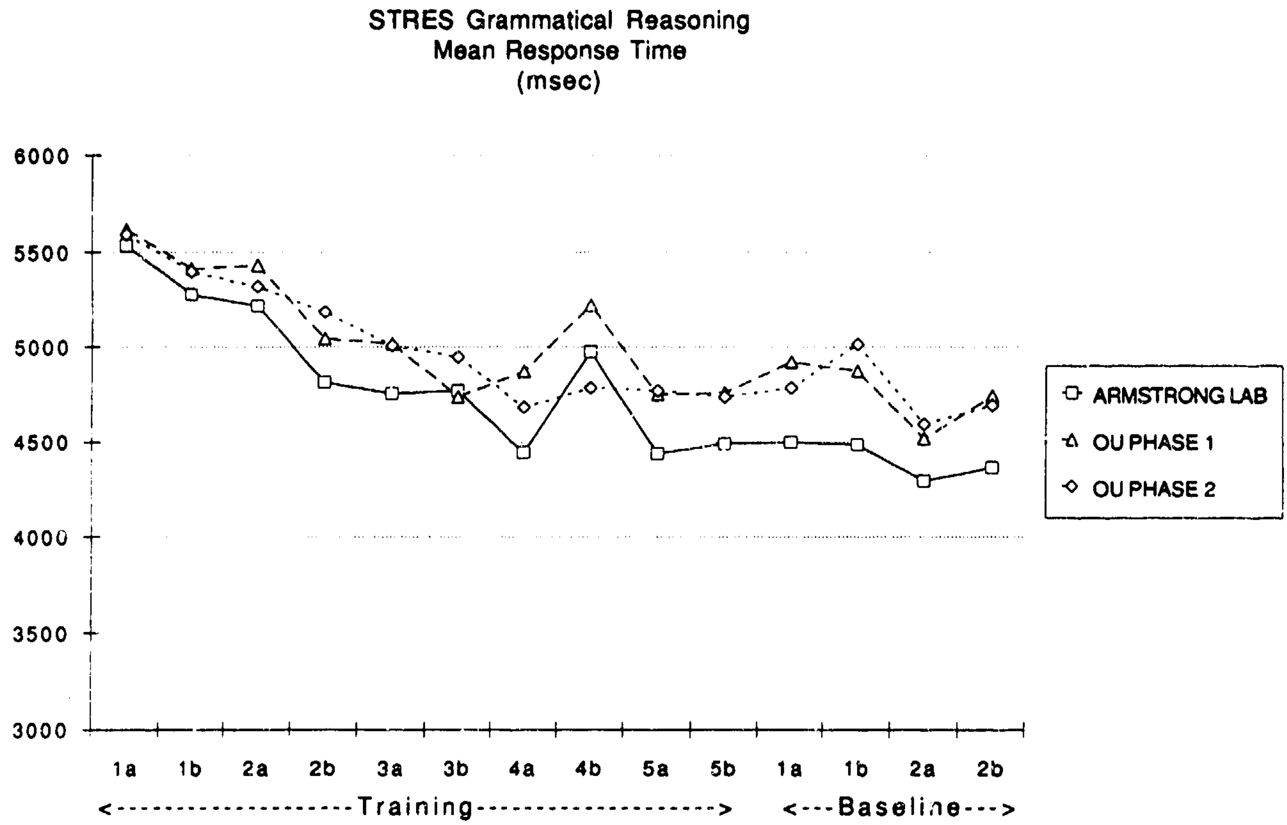
Unstable Tracking
Edge Violations
350
300
250
200
150
100
50
0
1
2
3
4
5
6
7
Trial
STRES
CTS '88
CTS '91

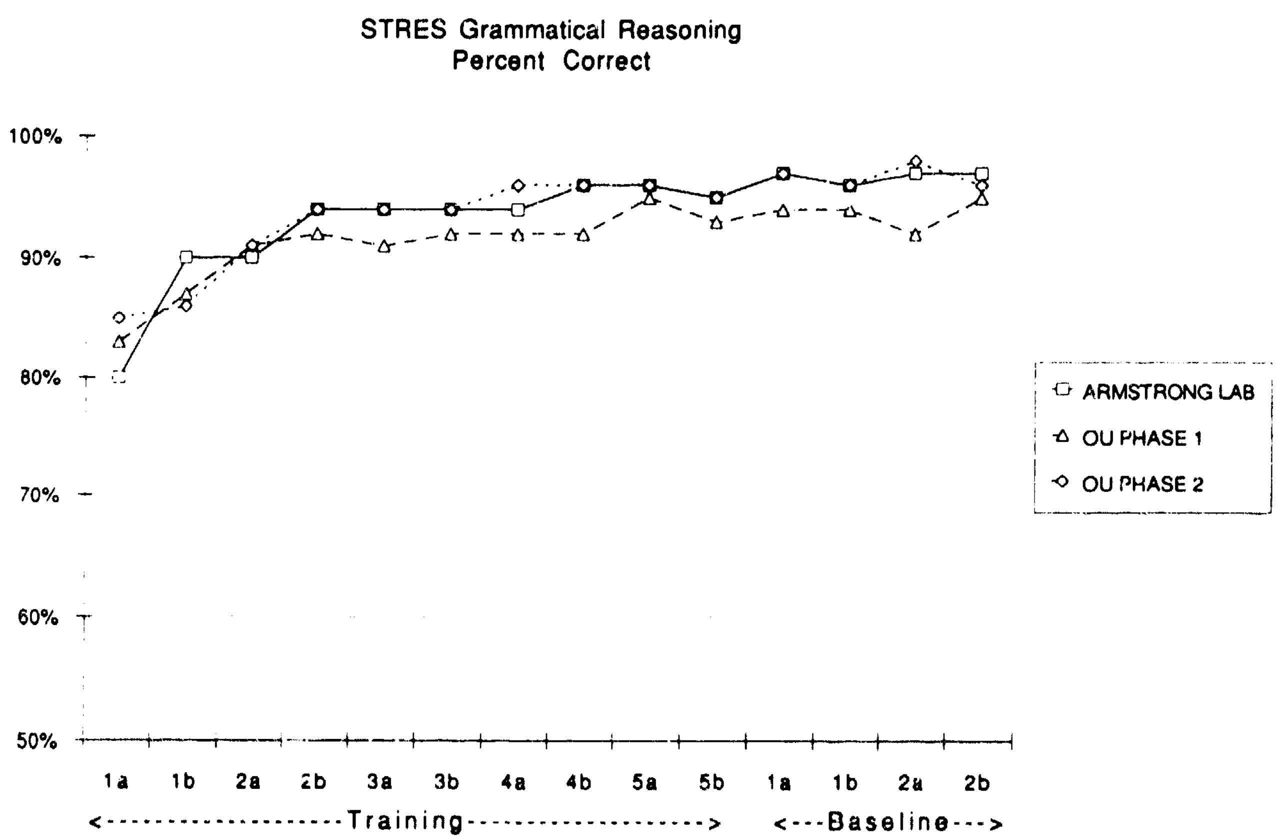


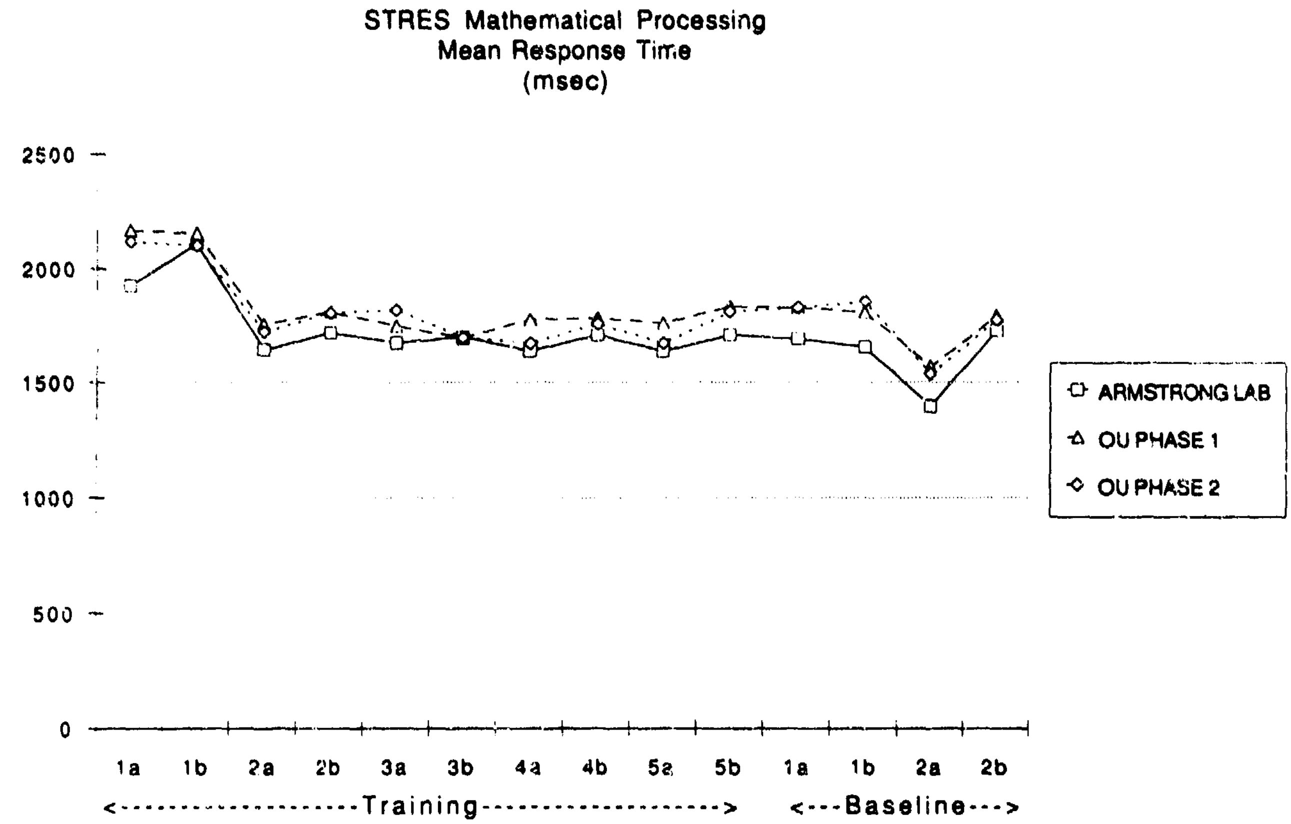
Unstable Tracking
RMS Error
45
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
Trial
STRES
CTS '88
CTS '91

APPENDIX I

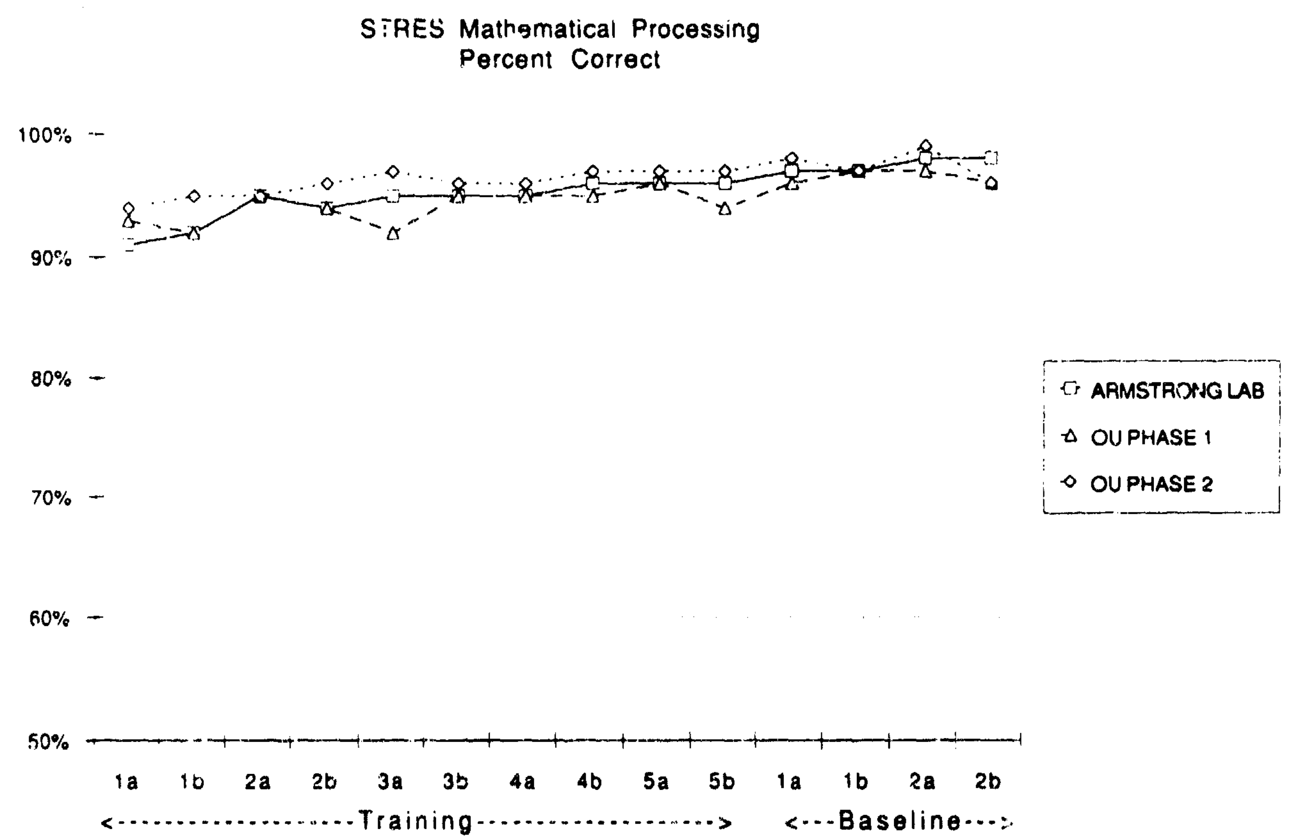
UNIVERSITY OF OKLAHOMA DATA
VS.
ARMSTRONG LABORATORY DATA


STRES Grammatical Reasoning
Mean Response Time
(msec)
6000
5500
5000
4500
4000
3500
3000
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2


STRES Grammatical Reasoning
Percent Correct
100%
90%
80%
70%
60%
50%
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

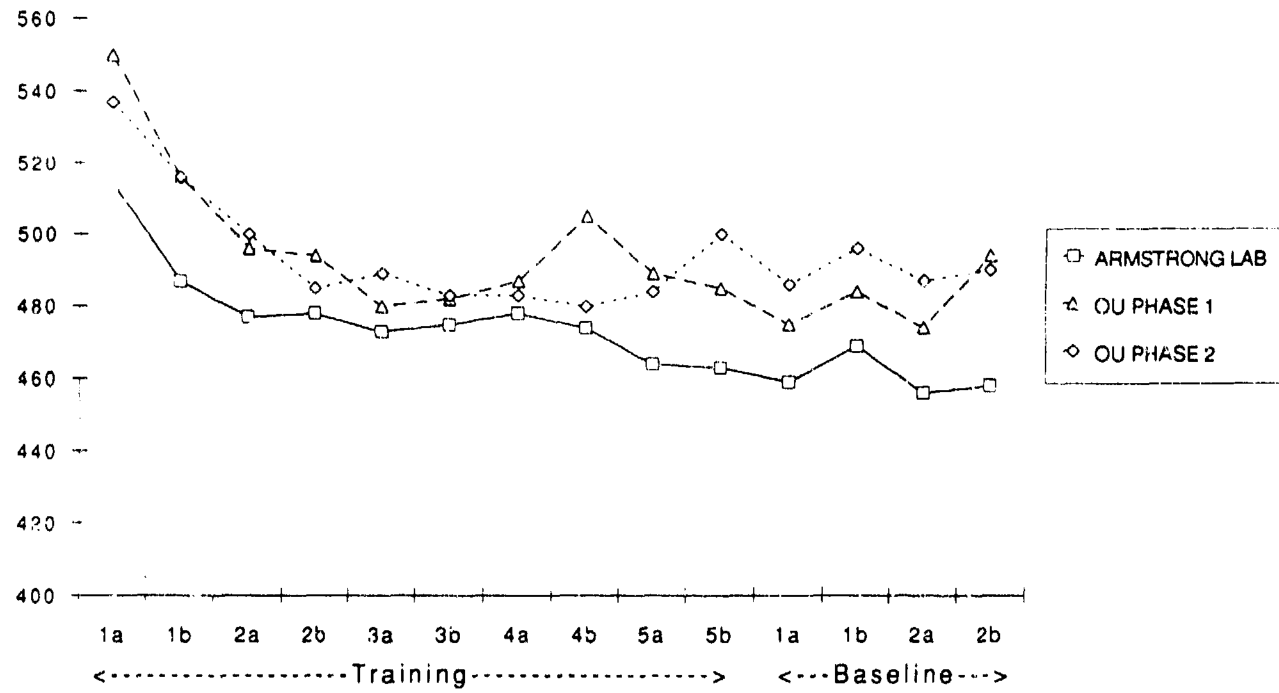


STRES Mathematical Processing
Mean Response Time
(msec)
2500
2000
1500
1000
500
0
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

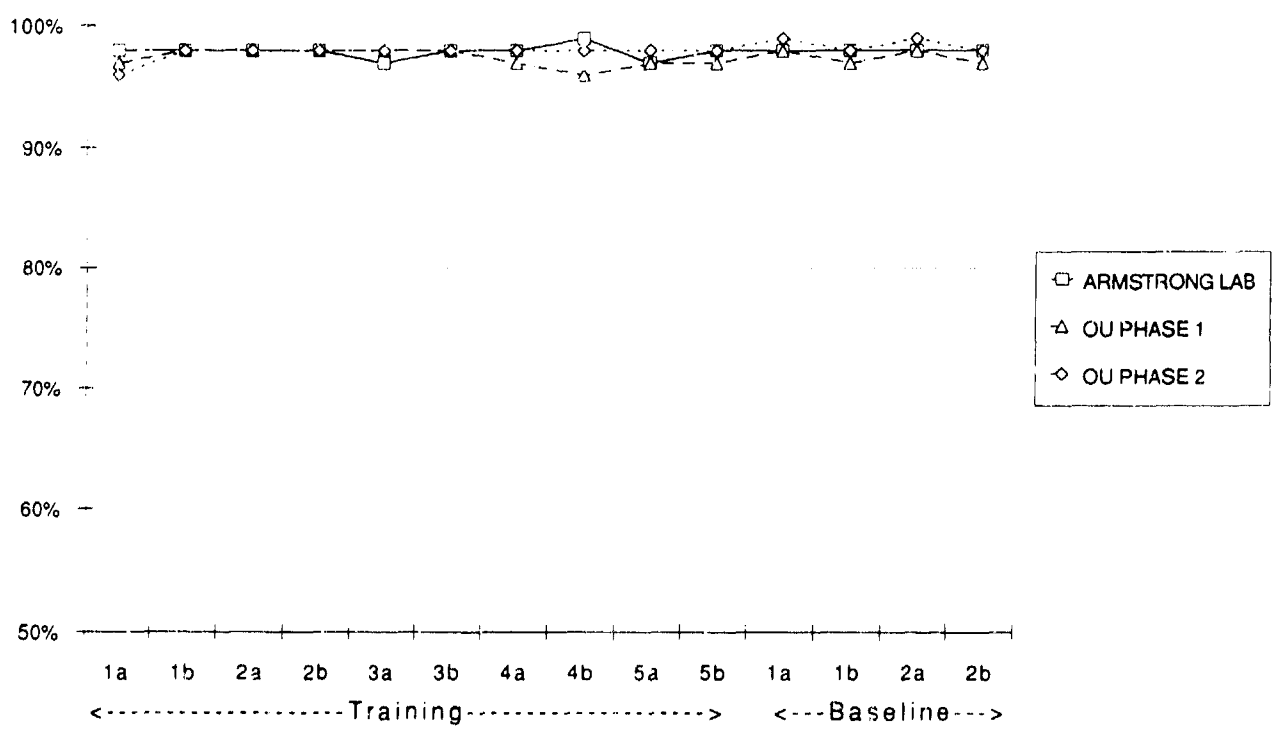


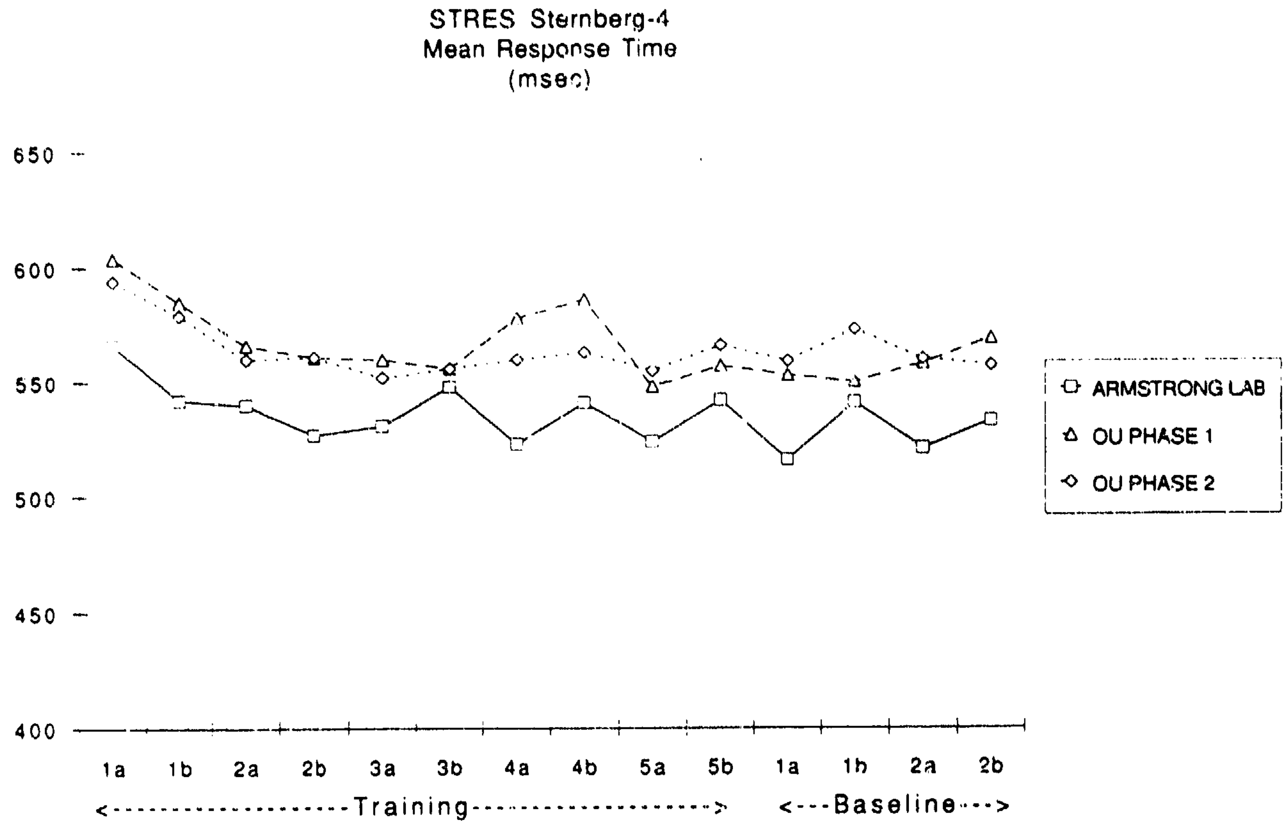
STRES Mathematical Processing
Percent Correct
100%
90%
80%
70%
60%
50%
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

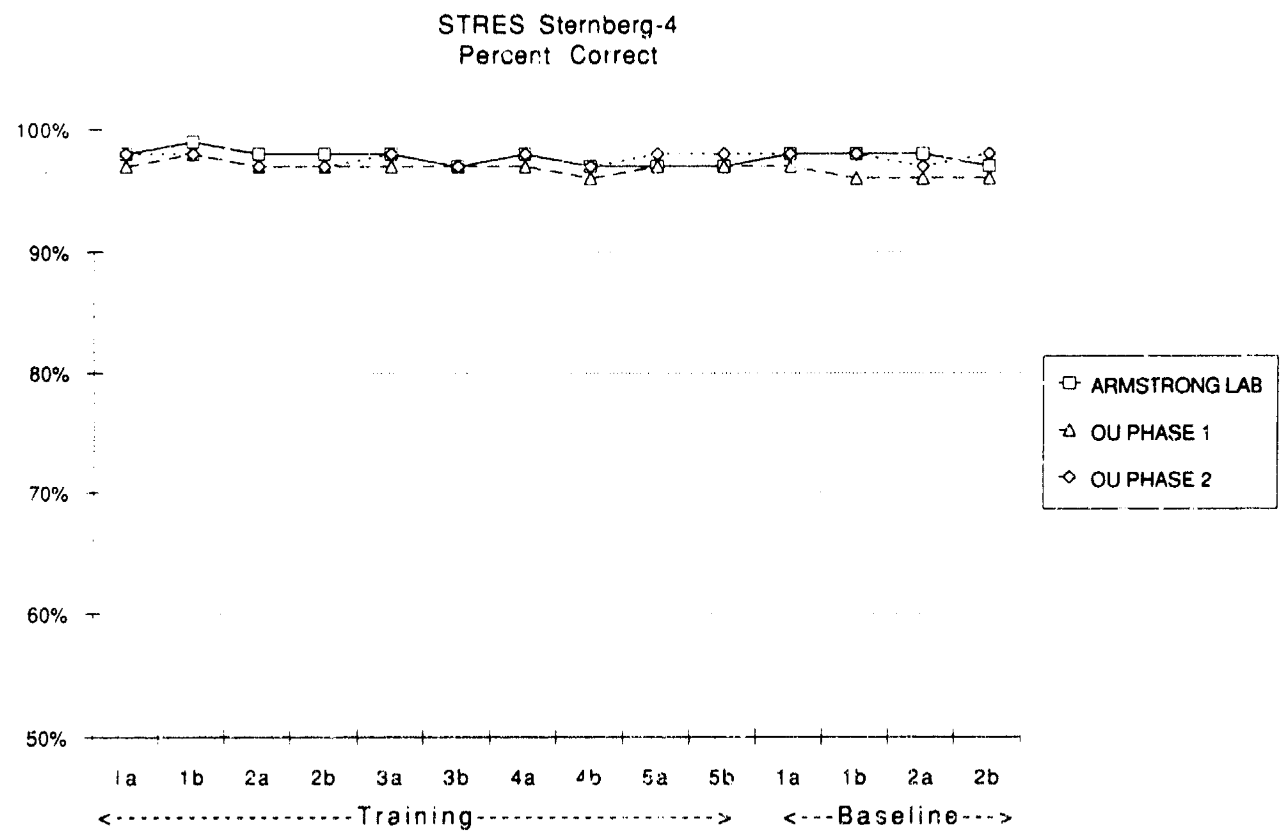
STRES Sternberg-2
Mean Response Time
(msec)

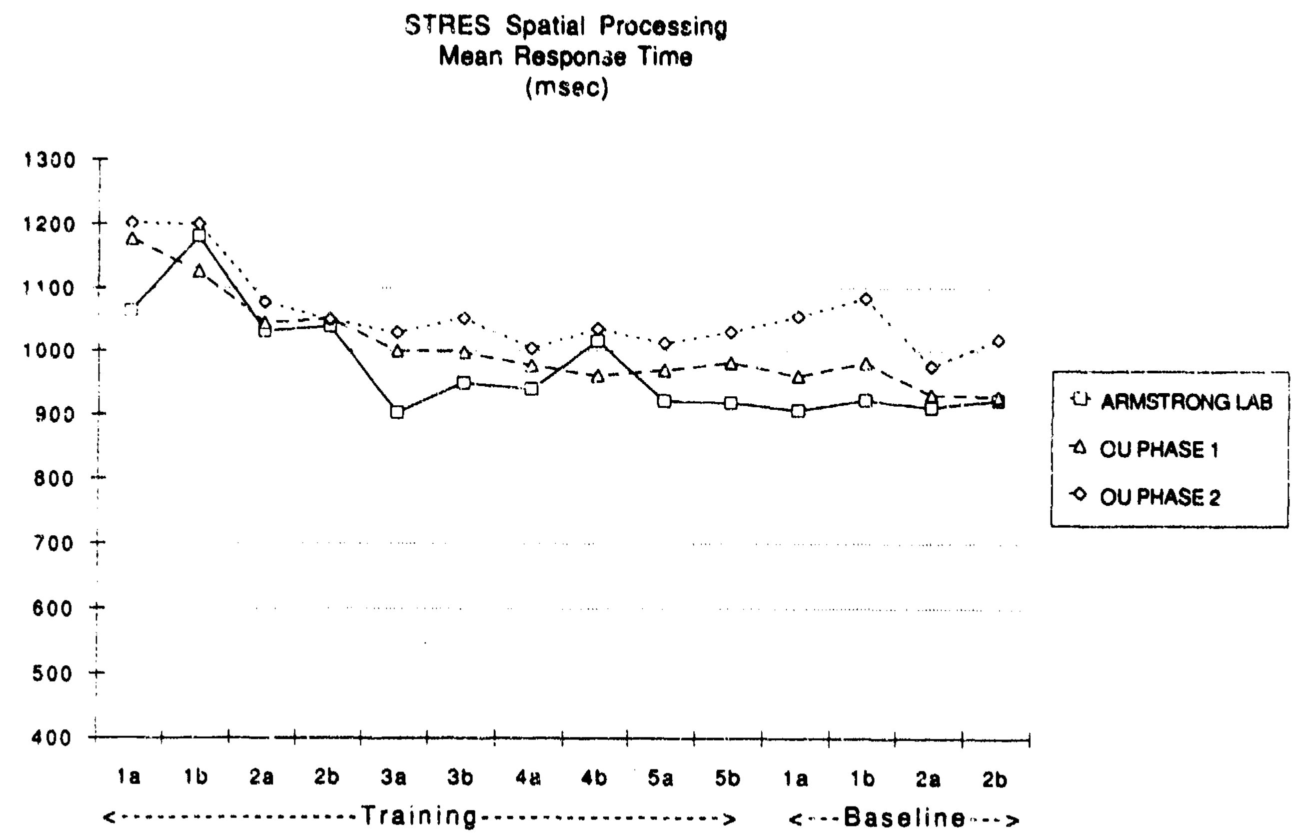


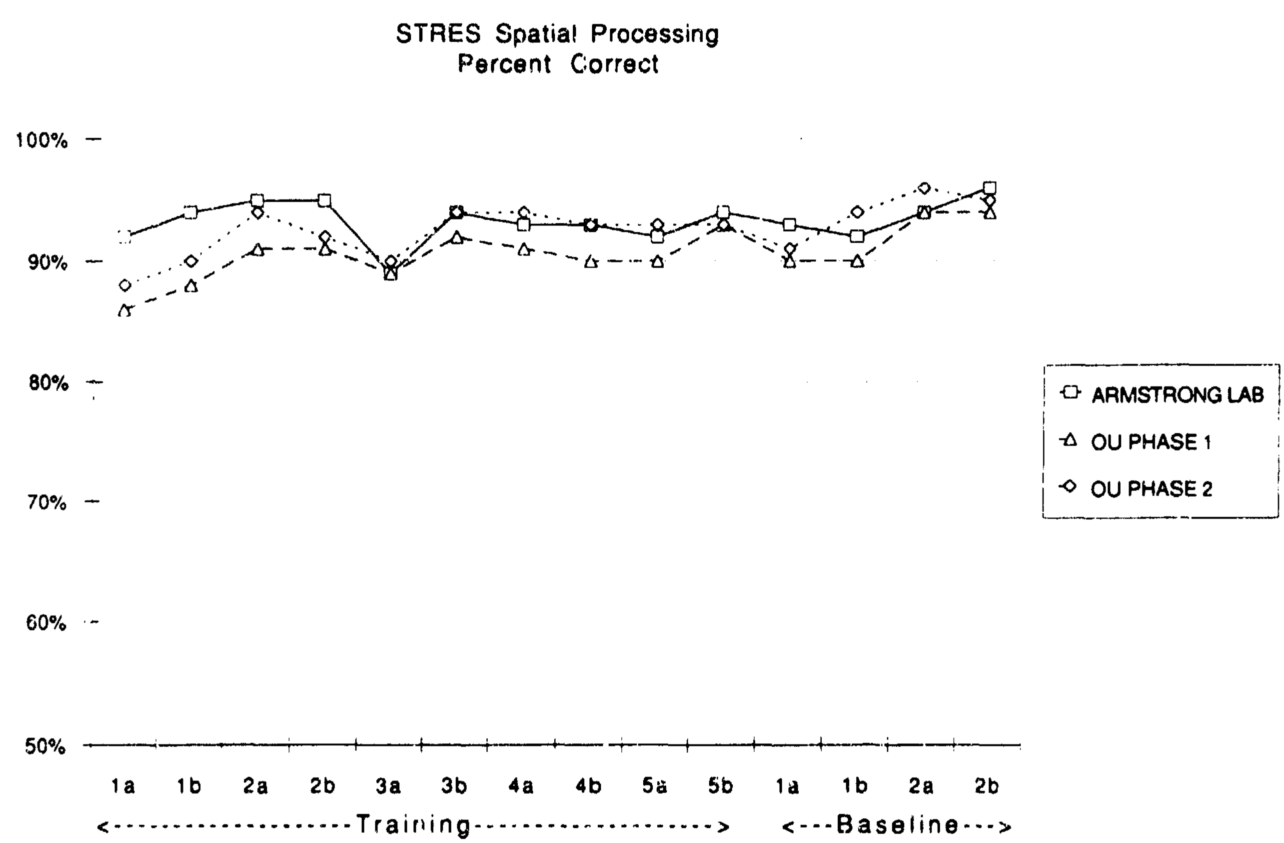
STRES Sternberg-2
Percent Correct

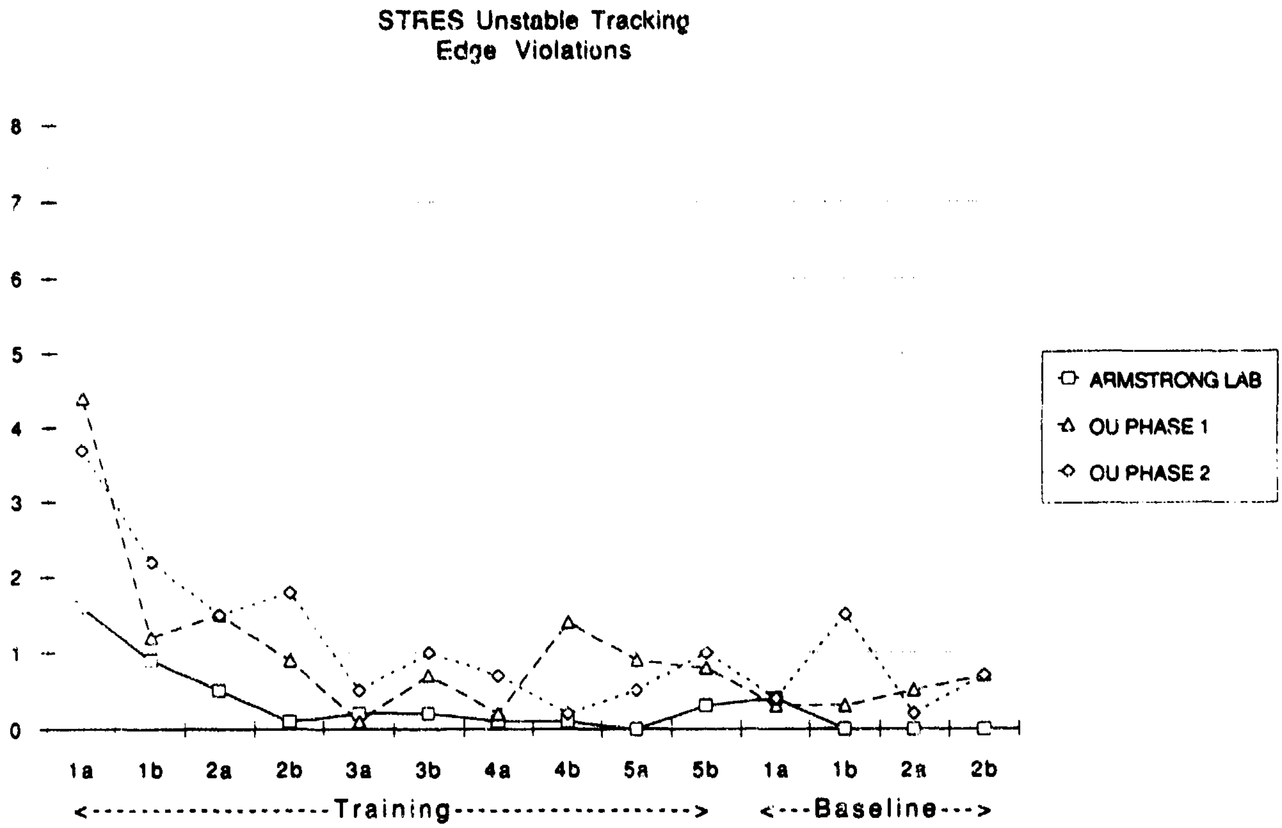


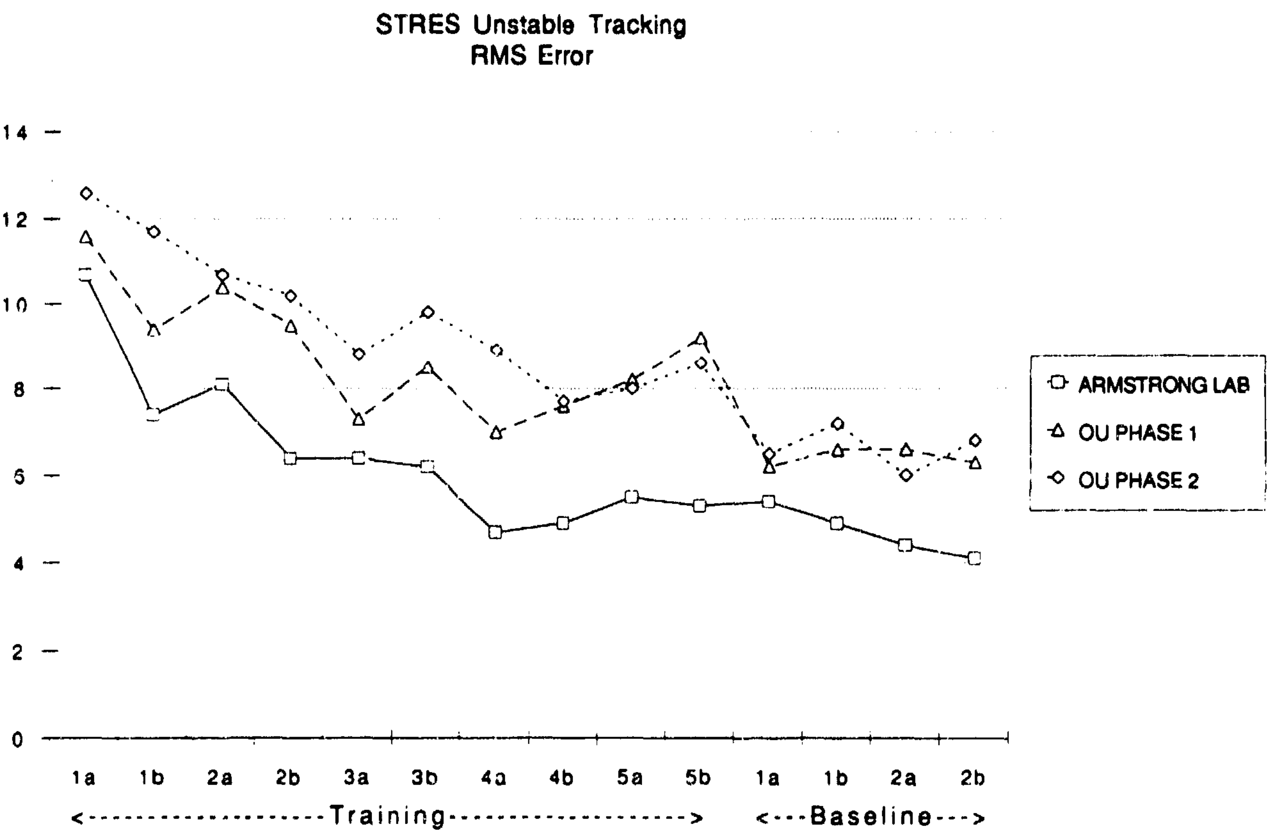

STRES Sternberg-4
Mean Response Time
(msec)
650
600
550
500
450
400
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<------------------Training------------------>
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

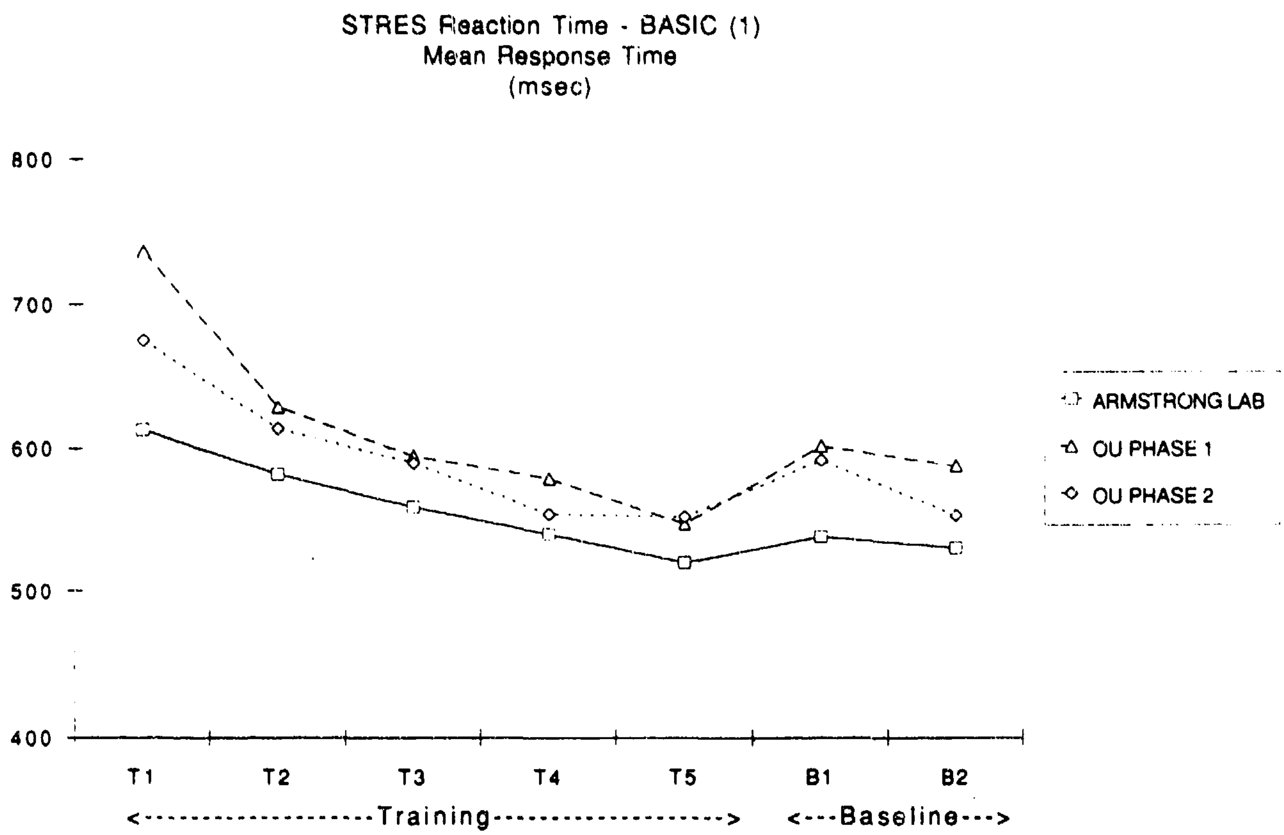

STRES Sternberg-4
Percent Correct
100%
90%
80%
70%
60%
50%
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<------------------Training------------------>
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

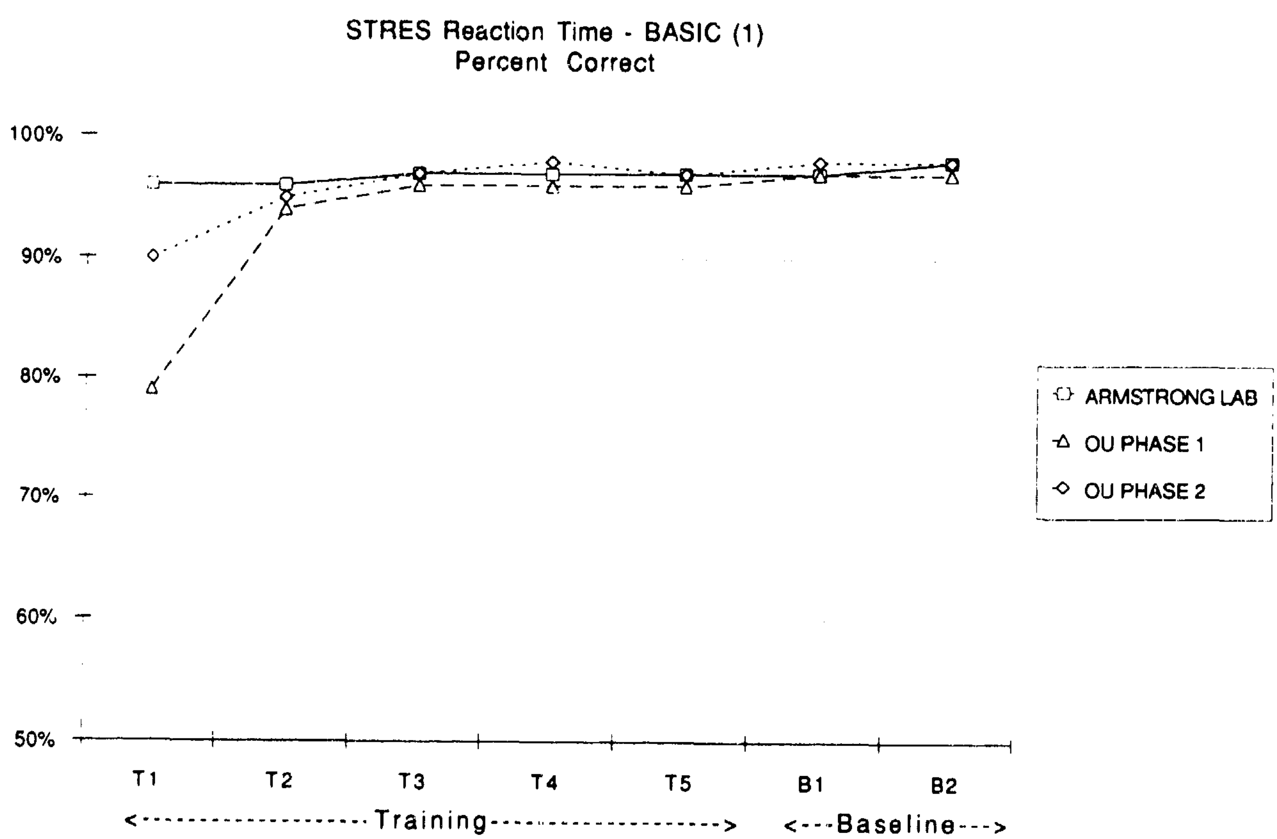

STRES Spatial Processing
Mean Response Time
(msec)
1300
1200
1100
1000
900
800
700
600
500
400
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

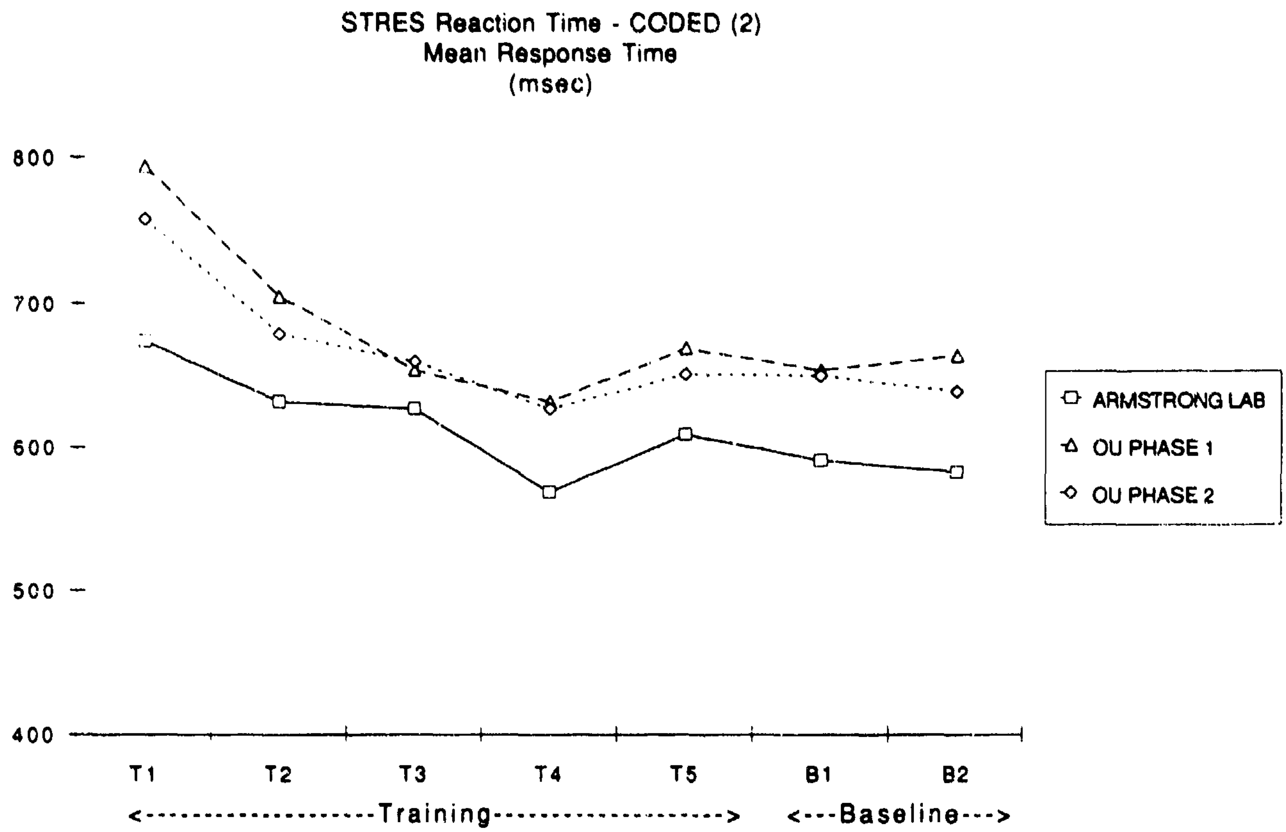

STRES Spatial Processing
Percent Correct
100%
90%
80%
70%
60%
50%
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

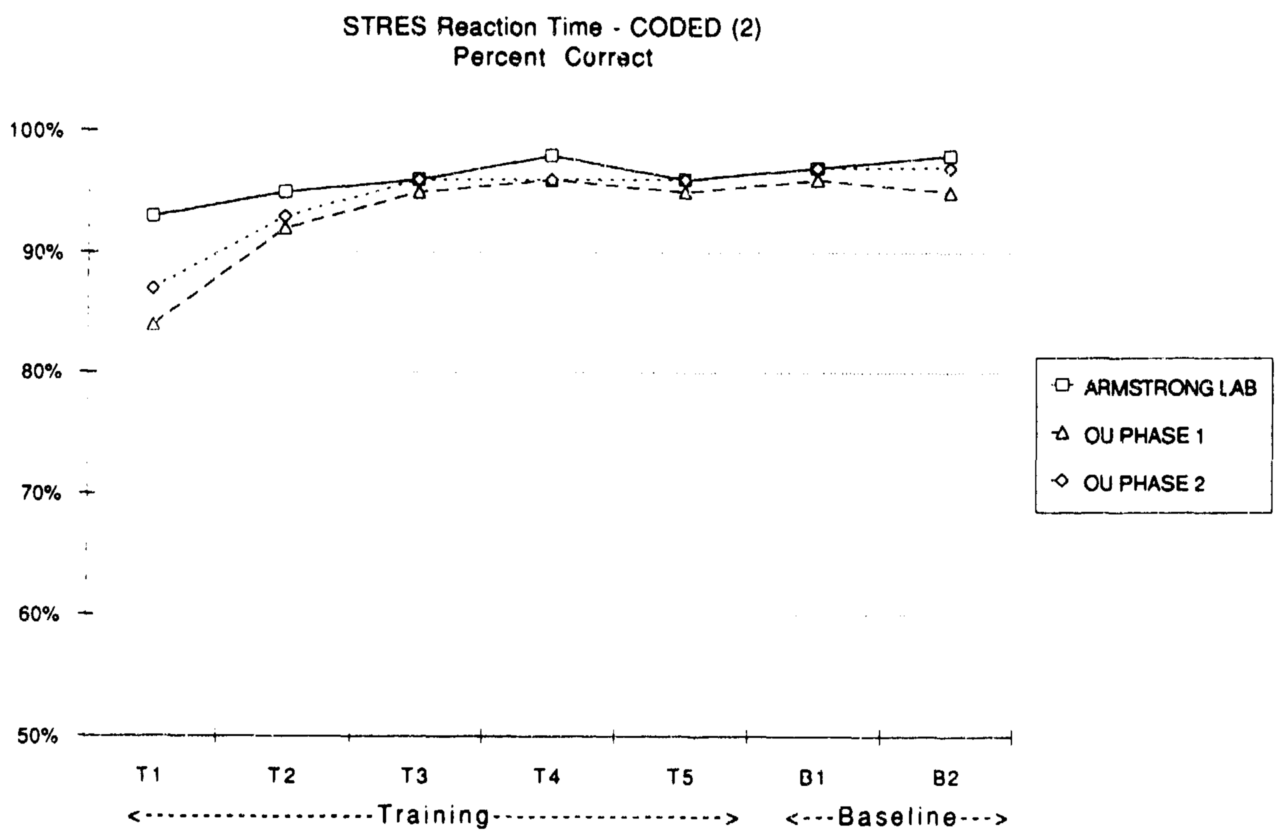

STRES Unstable Tracking
Edge Violations
8
7
6
5
4
3
2
1
0
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<------------------Training------------------> <---Baseline--->

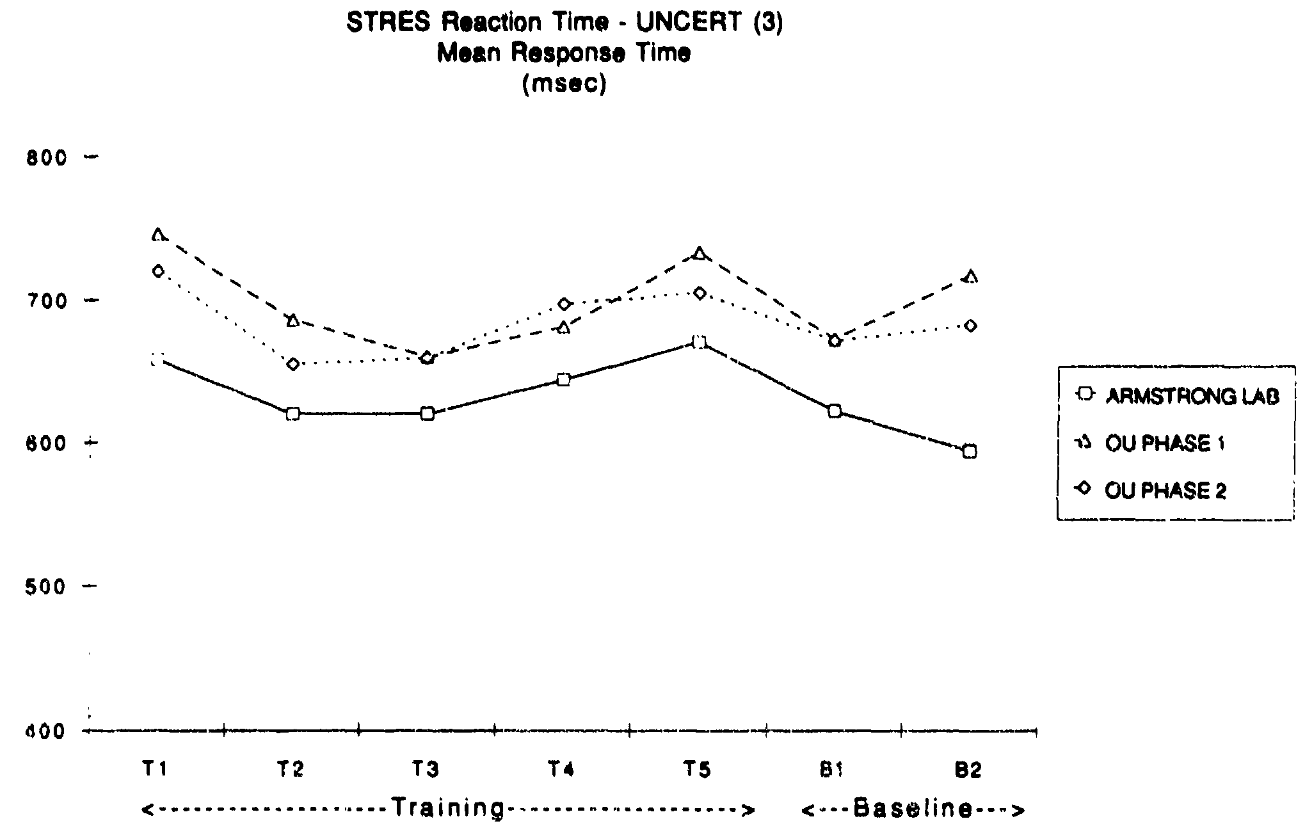

STRES Unstable Tracking
RMS Error
14
12
10
8
6
4
2
0
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<------------------Training------------------> <---Baseline--->

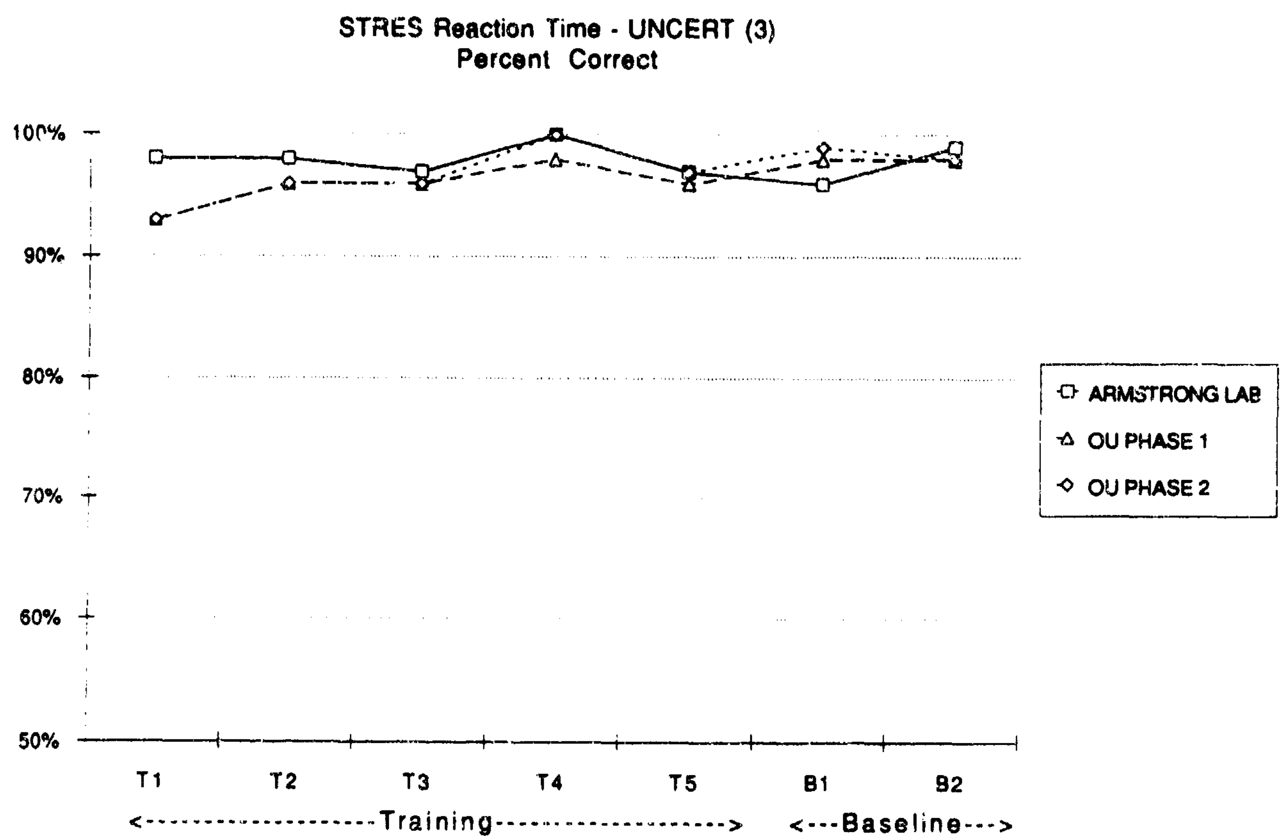

STRES Reaction Time - BASIC (1)
Mean Response Time
(msec)
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2


STRES Reaction Time - BASIC (1)
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

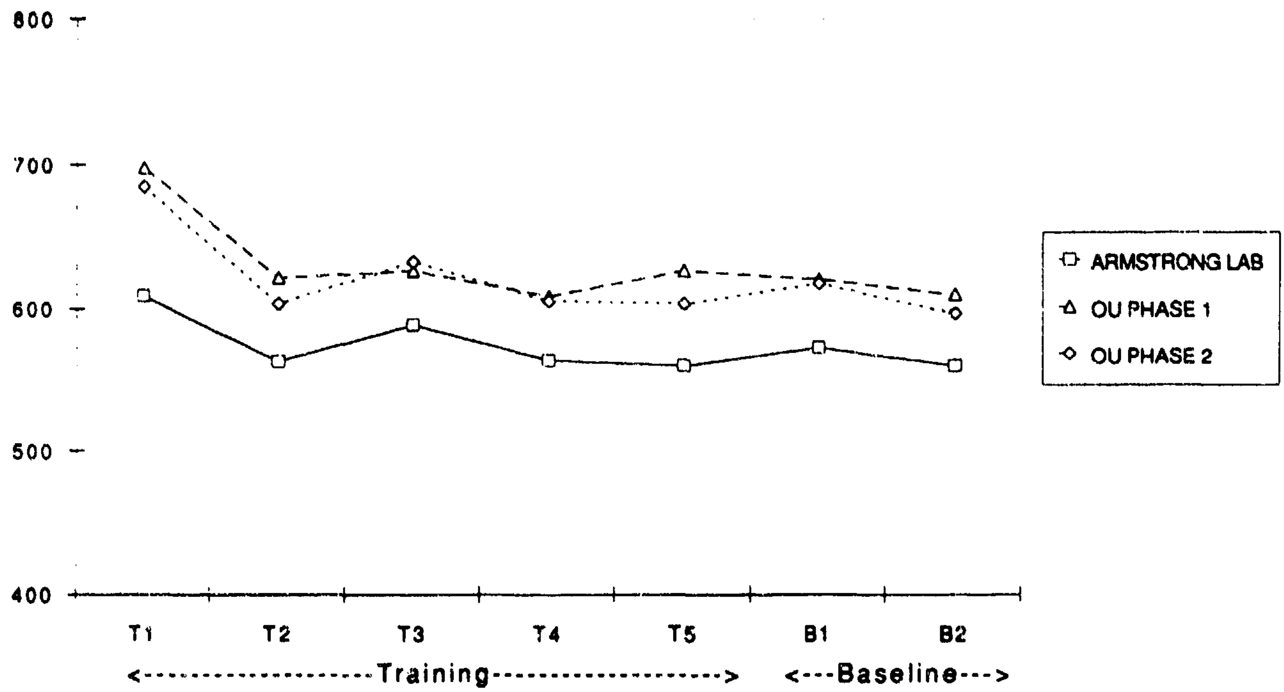

STRES Reaction Time - CODED (2)
Mean Response Time
(msec)
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2


STRES Reaction Time - CODED (2)
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

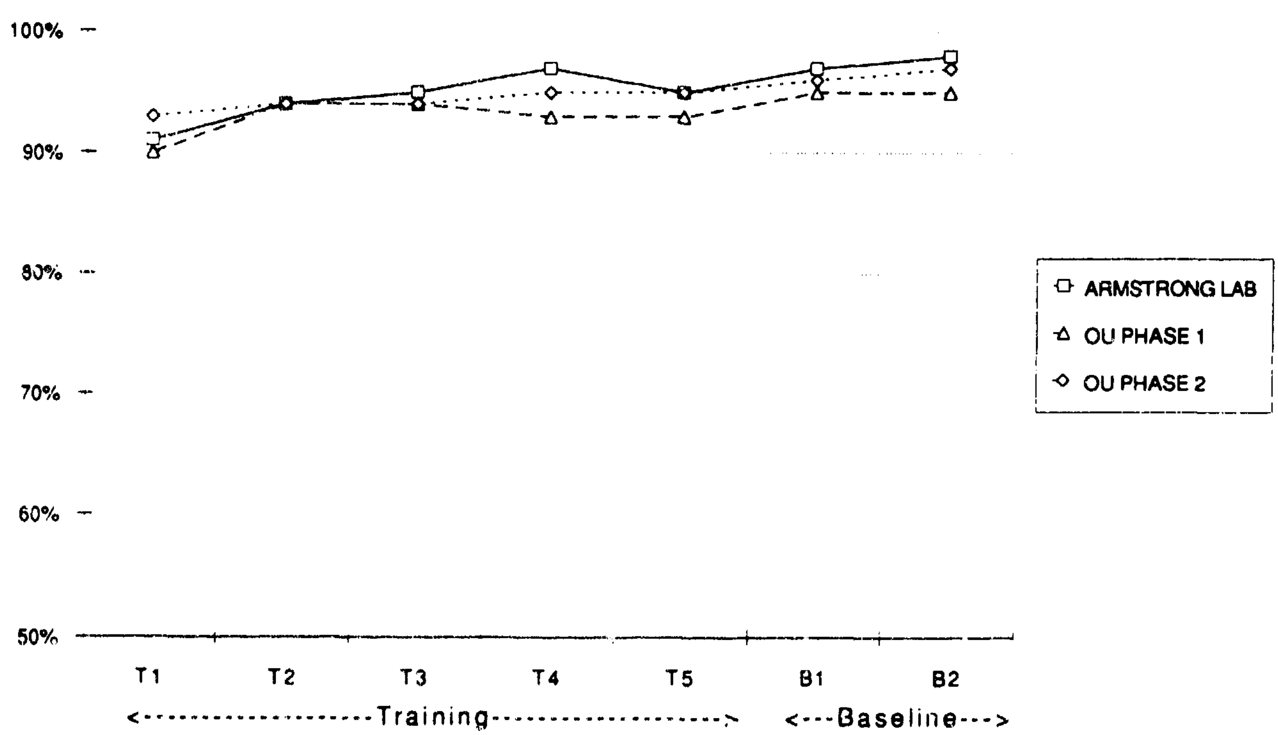

STRES Reaction Time - UNCERT (3)
Mean Response Time
(msec)
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

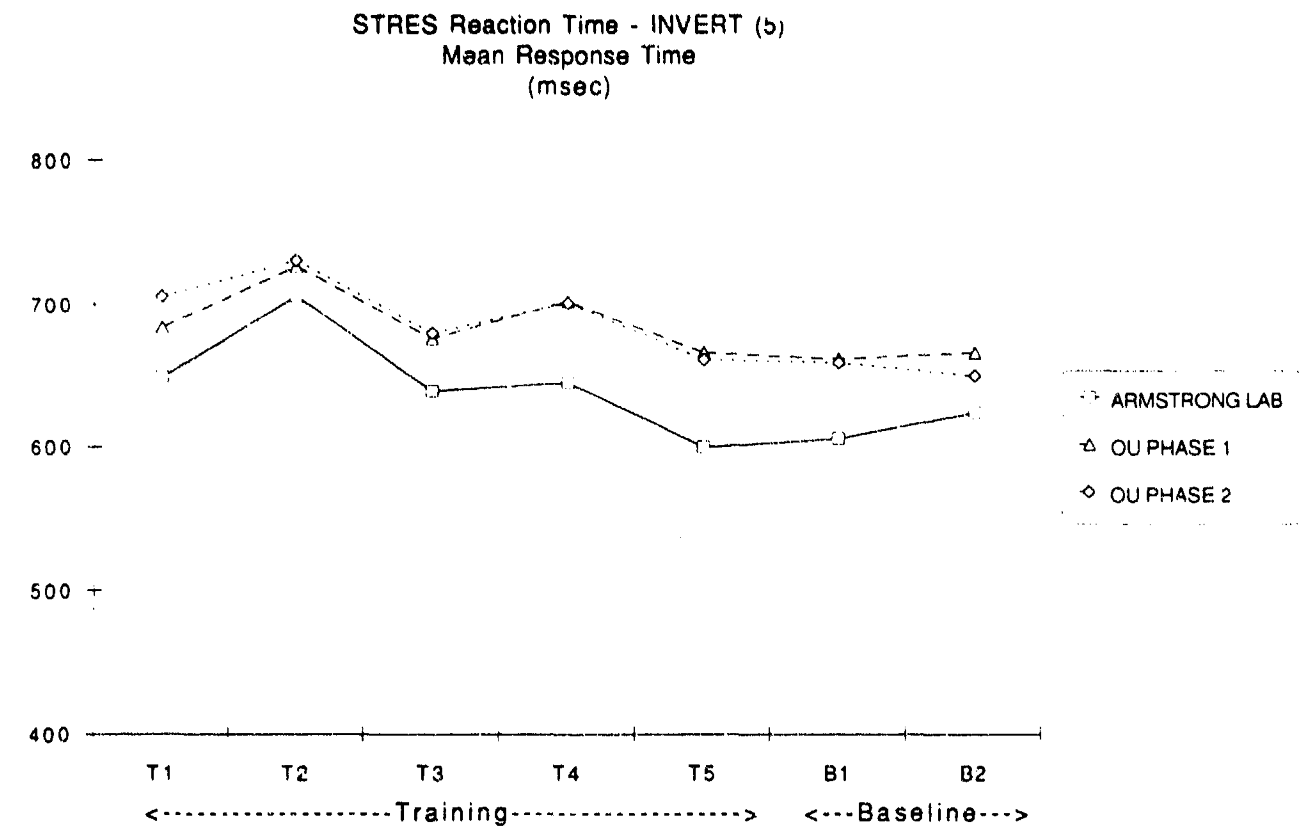

STRES Reaction Time - UNCERT (3)
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

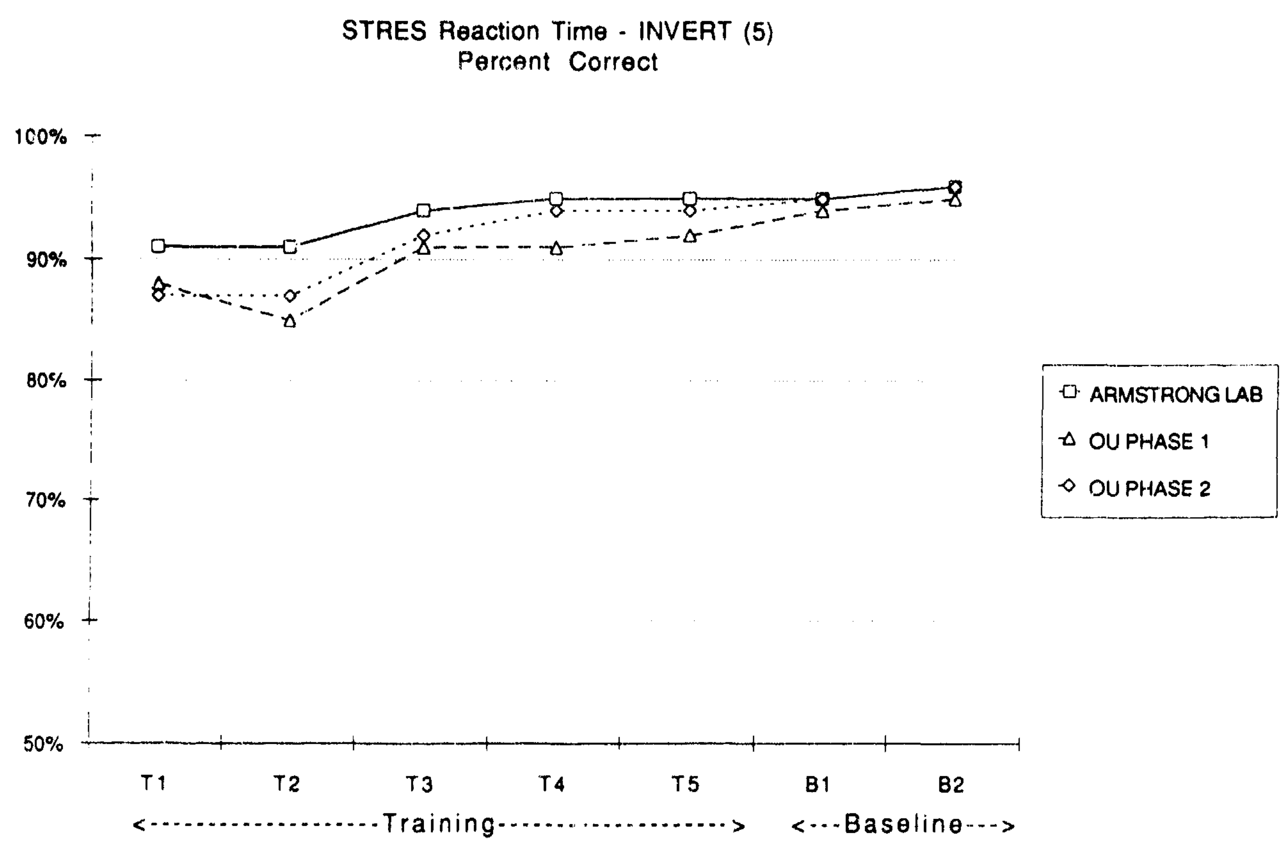
STRES Reaction Time - DOUBLE (4)
Mean Response Time
(msec)

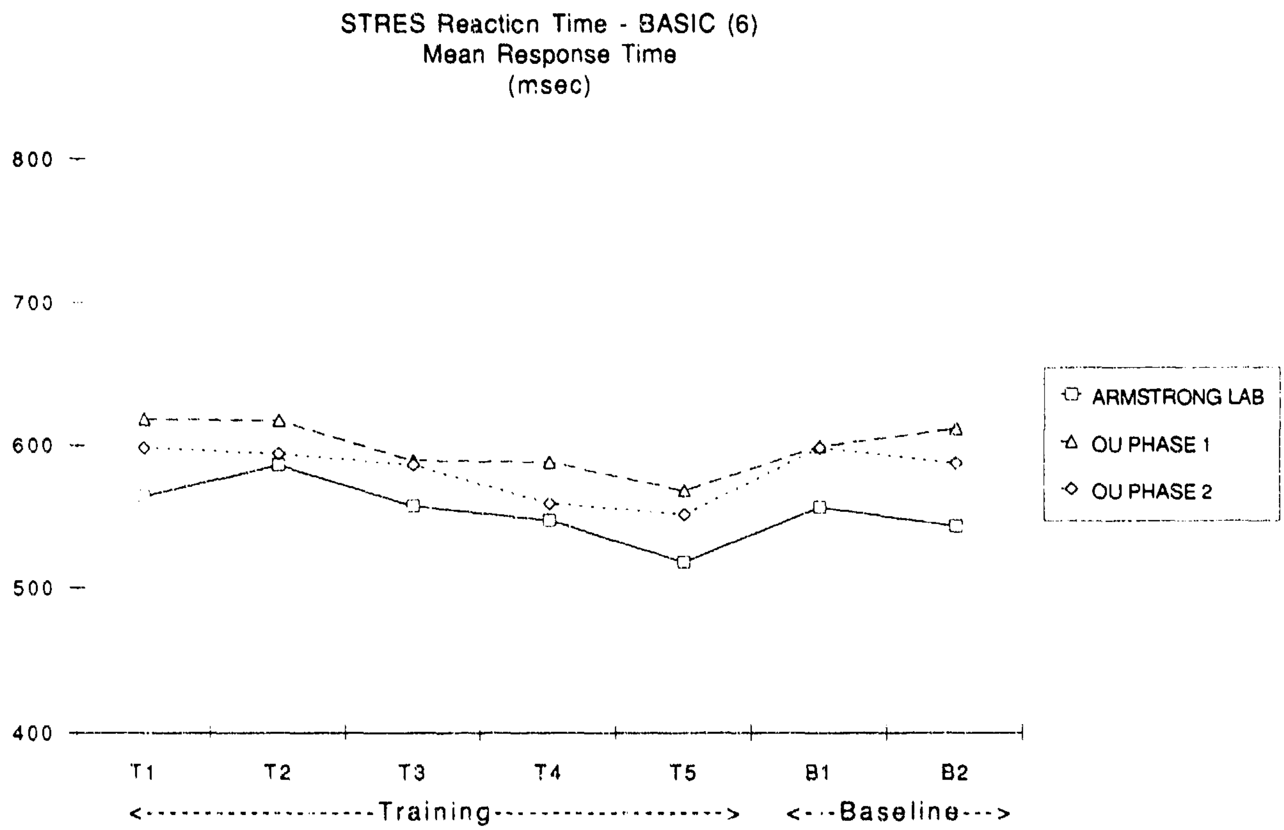


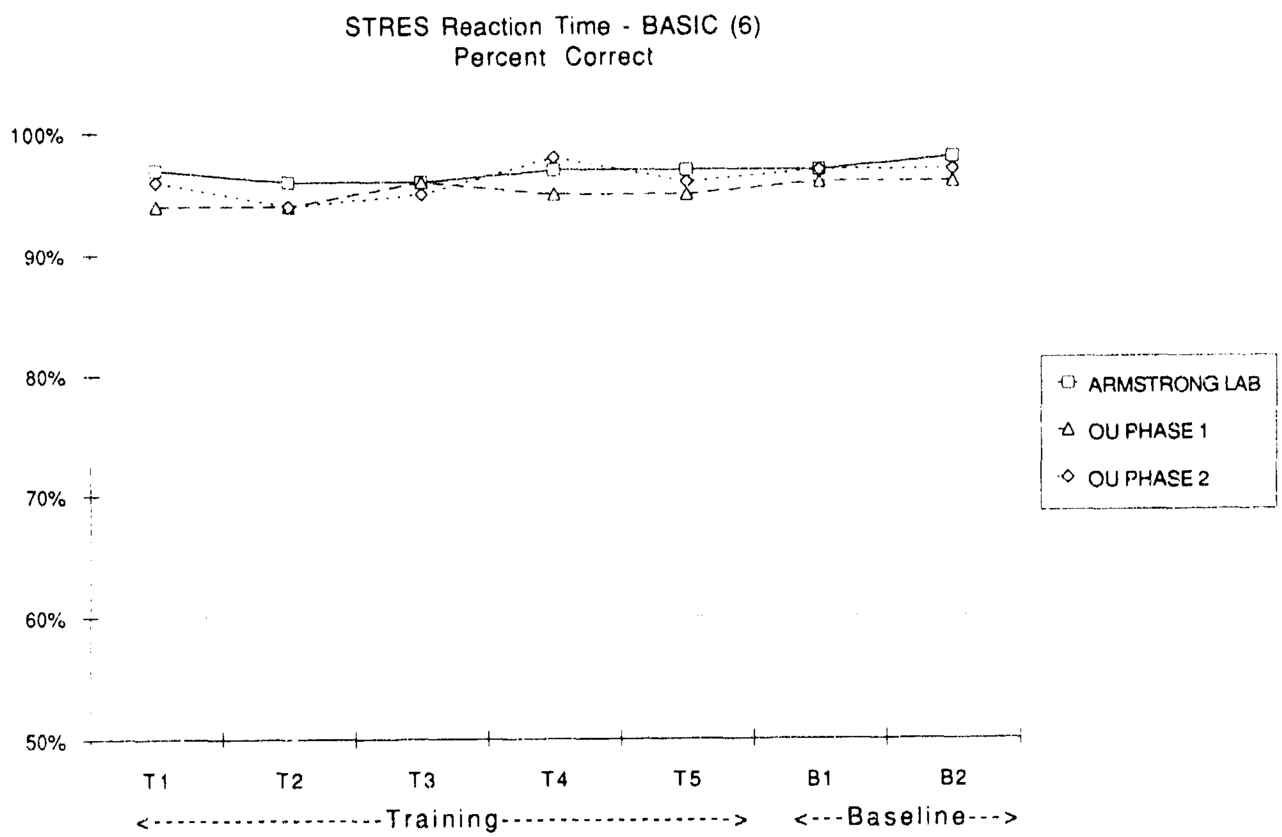
STRES Reaction Time - DOUBLE (4)
Percent Correct

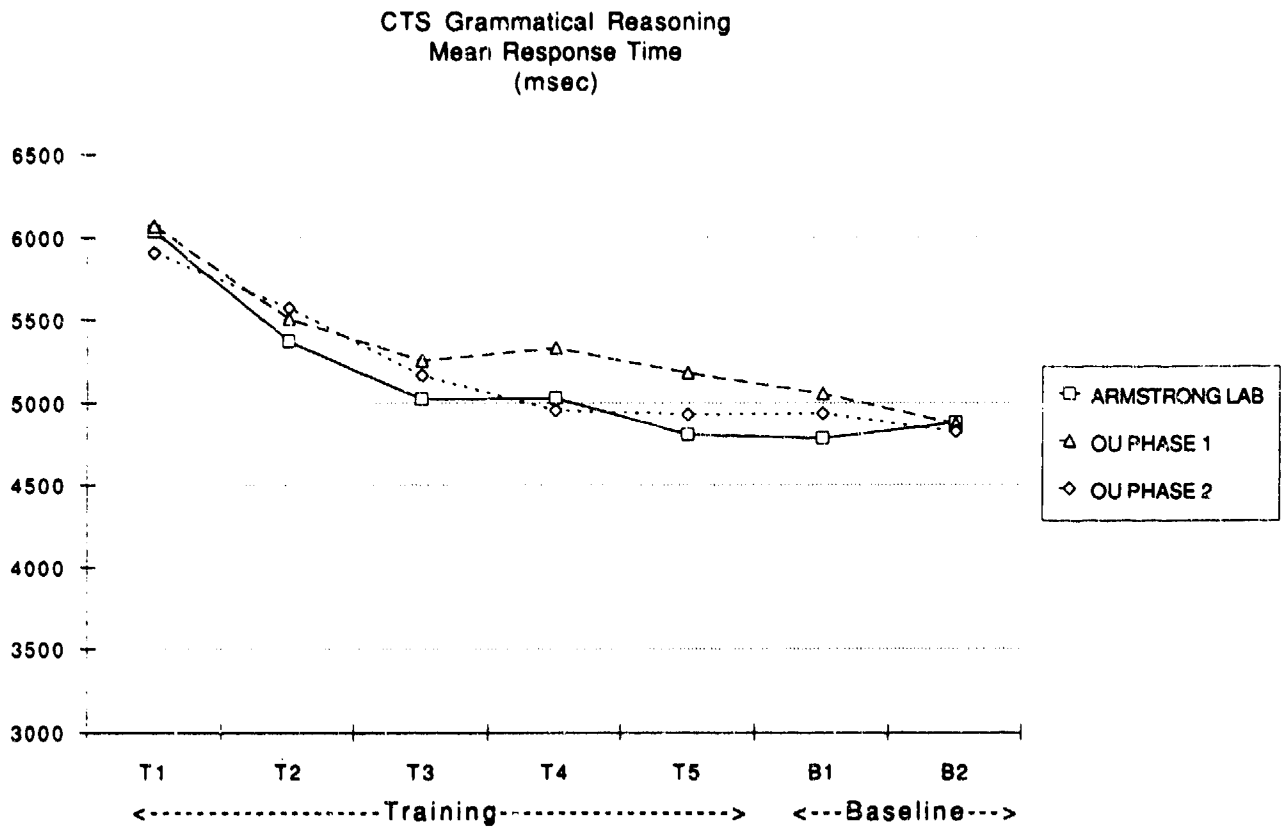


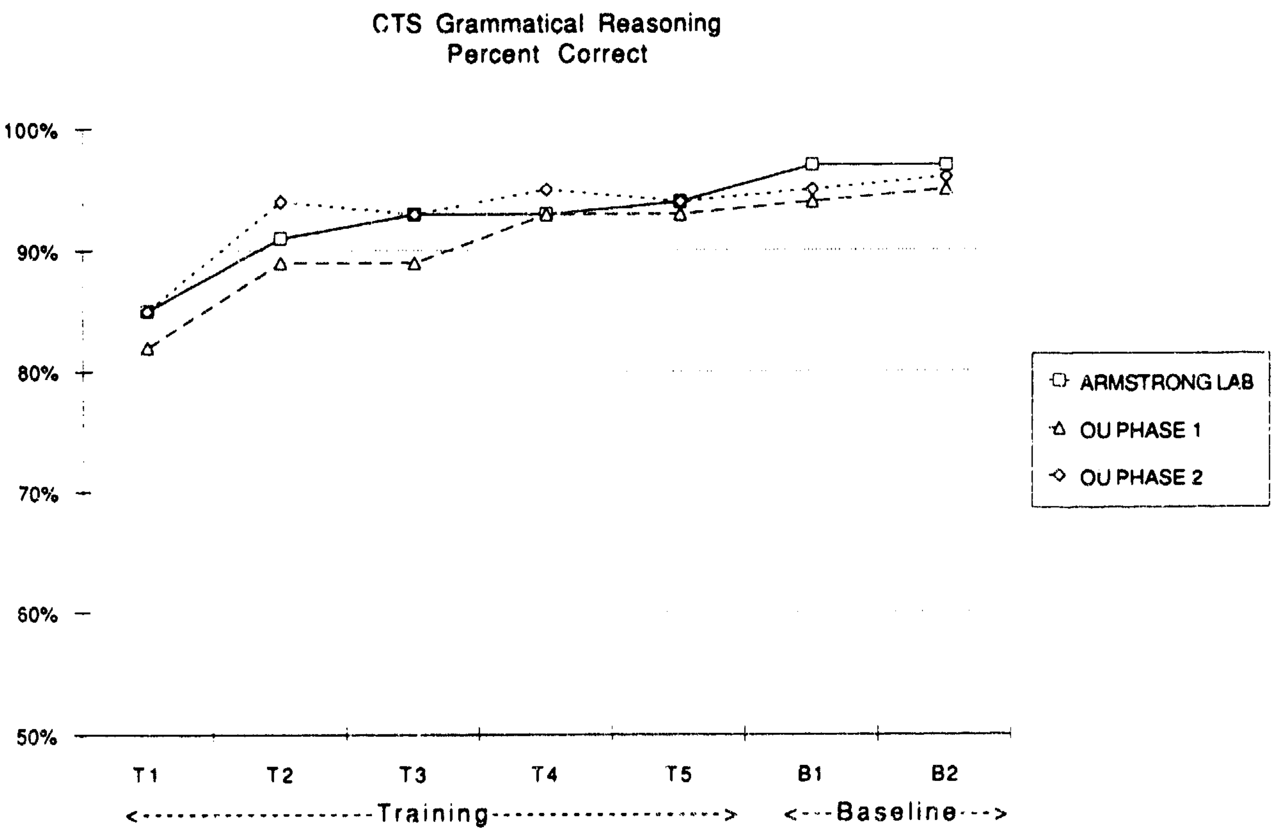

STRES Reaction Time - INVERT (5)
Mean Response Time
(msec)
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2


STRES Reaction Time - INVERT (5)
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

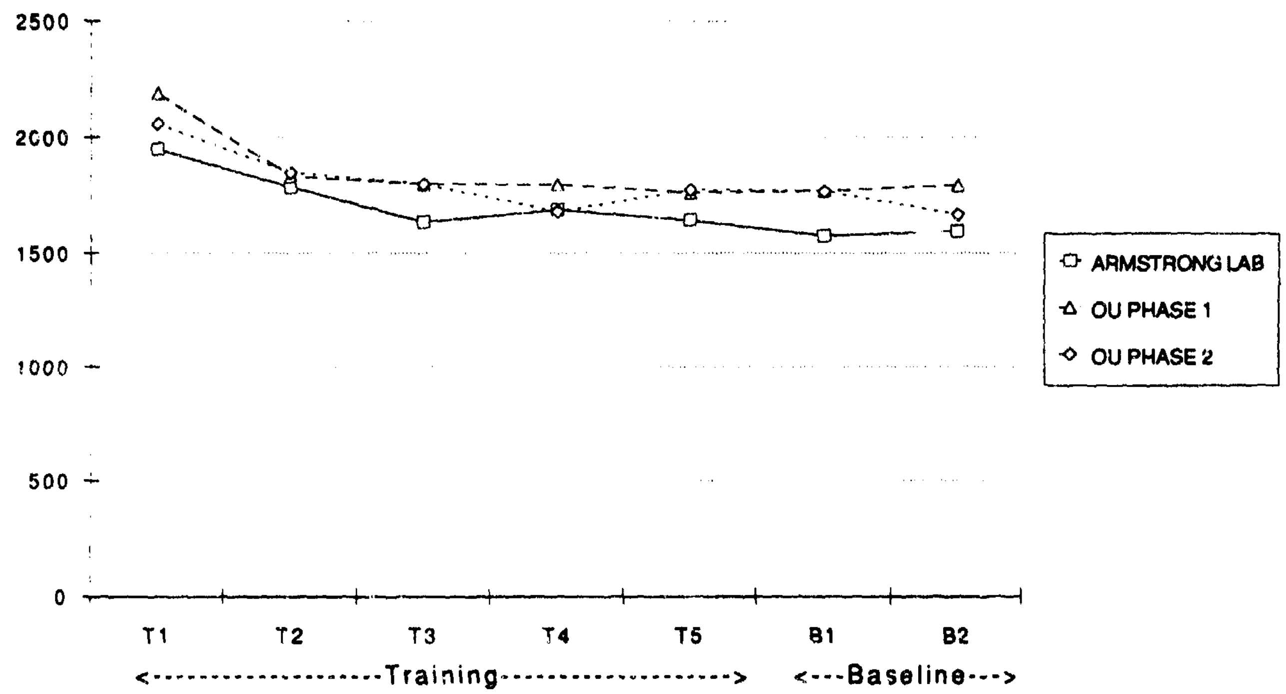

STRES Reaction Time - BASIC (6)
Mean Response Time
(msec)
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------> <---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2


STRES Reaction Time - BASIC (6)
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------> <---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

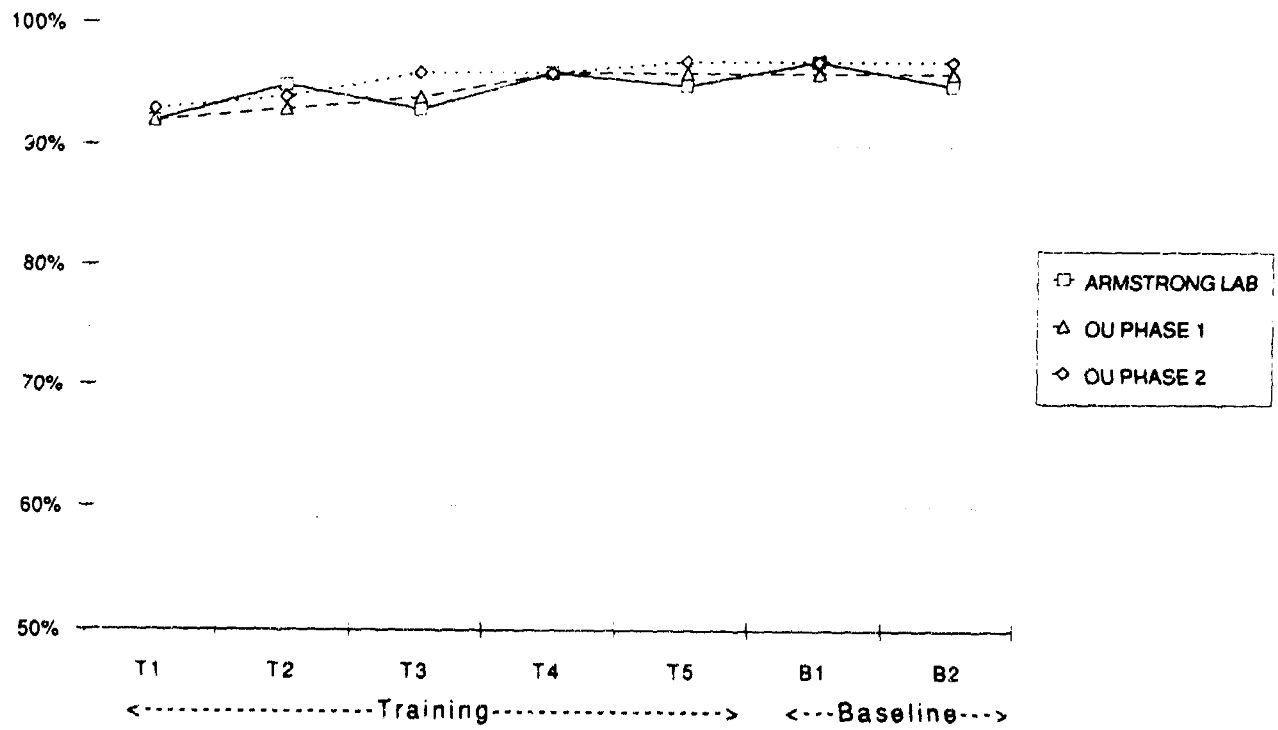

CTS Grammatical Reasoning
Mean Response Time
(msec)
6500
6000
5500
5000
4500
4000
3500
3000
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

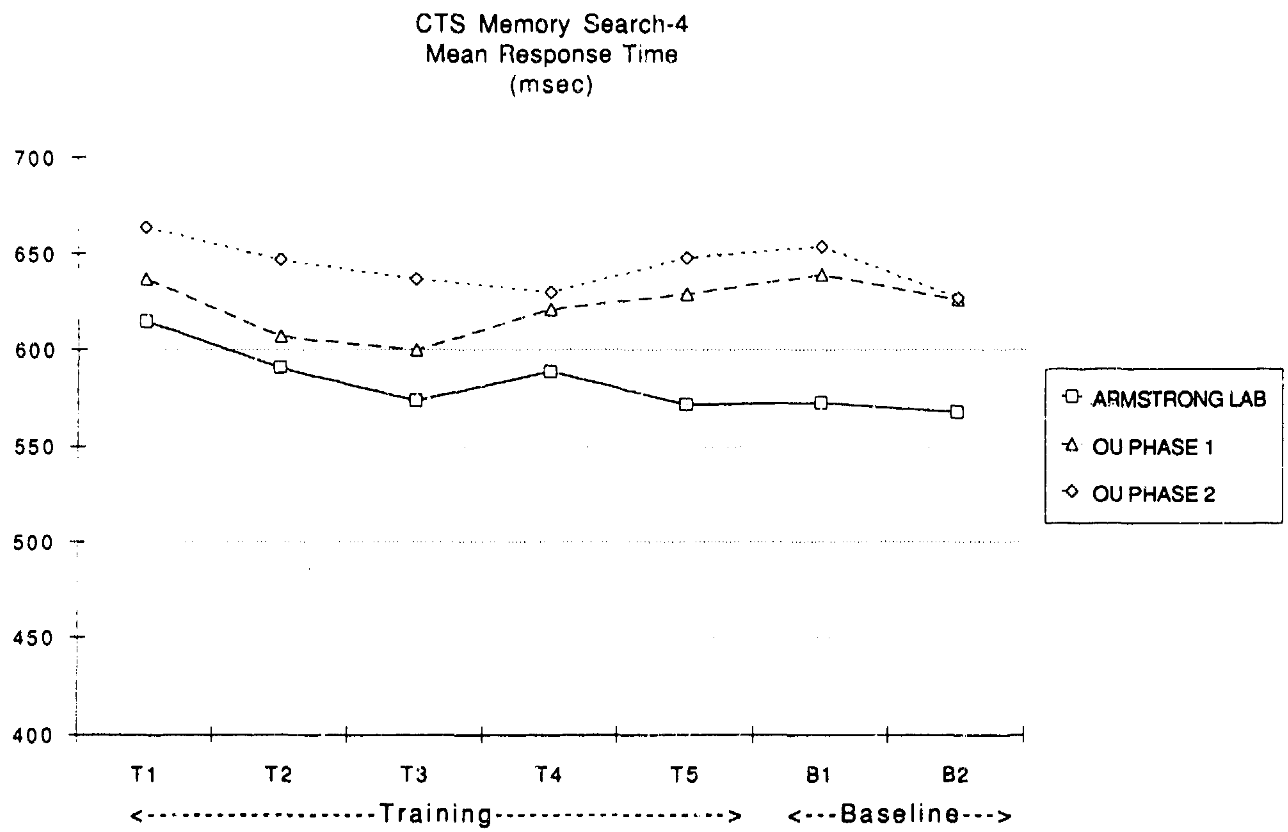

CTS Grammatical Reasoning
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

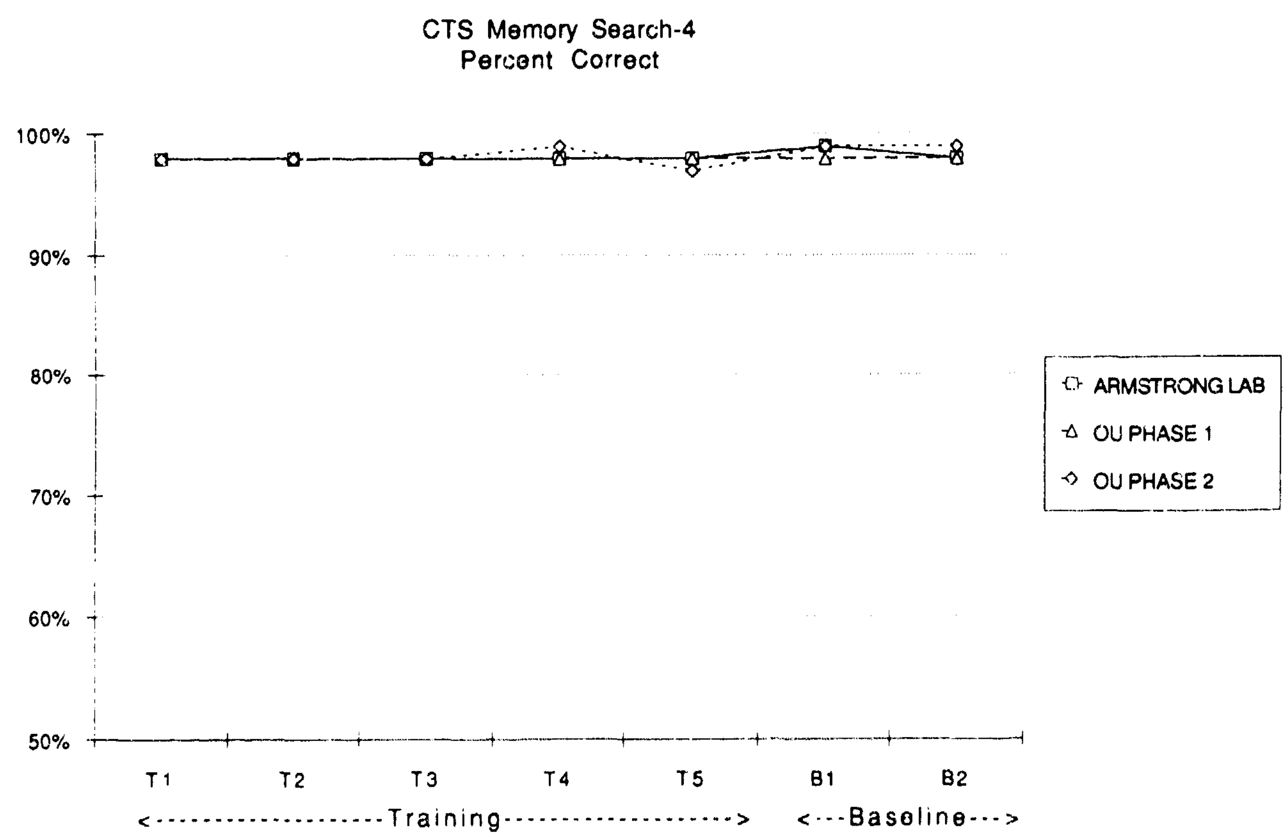
CTS Mathematical Processing
Mean Response Time
(msec)

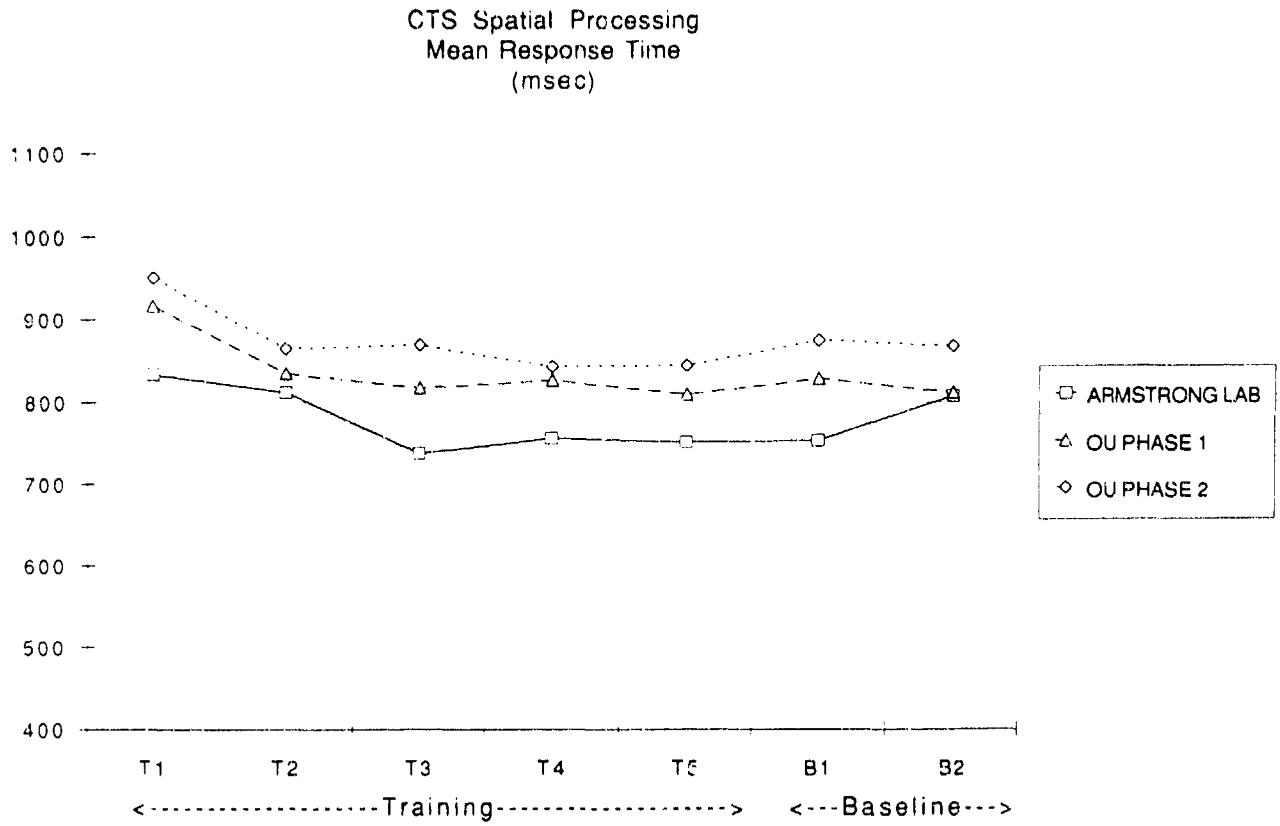


CTS Mathematical Processing
Percent Correct

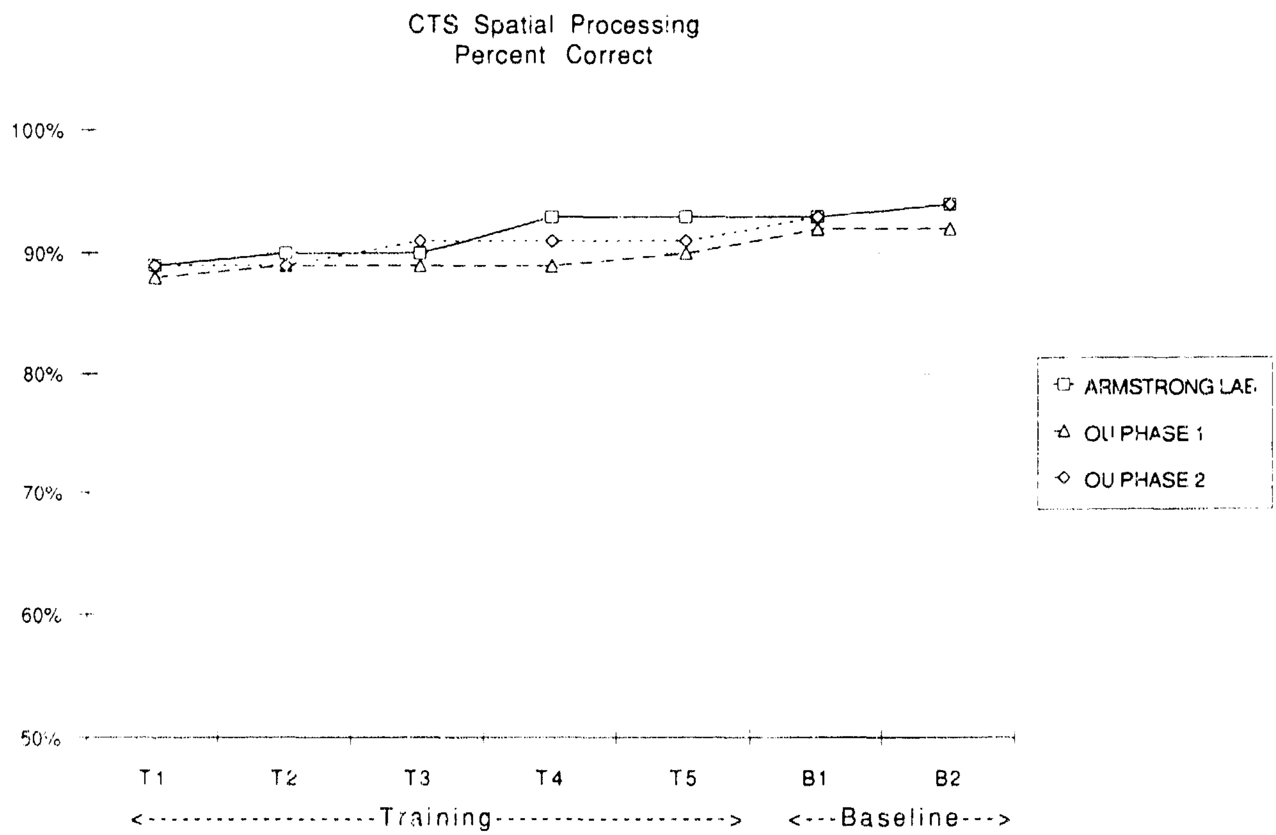



CTS Memory Search-4
Mean Response Time
(msec)
700
650
600
550
500
450
400
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

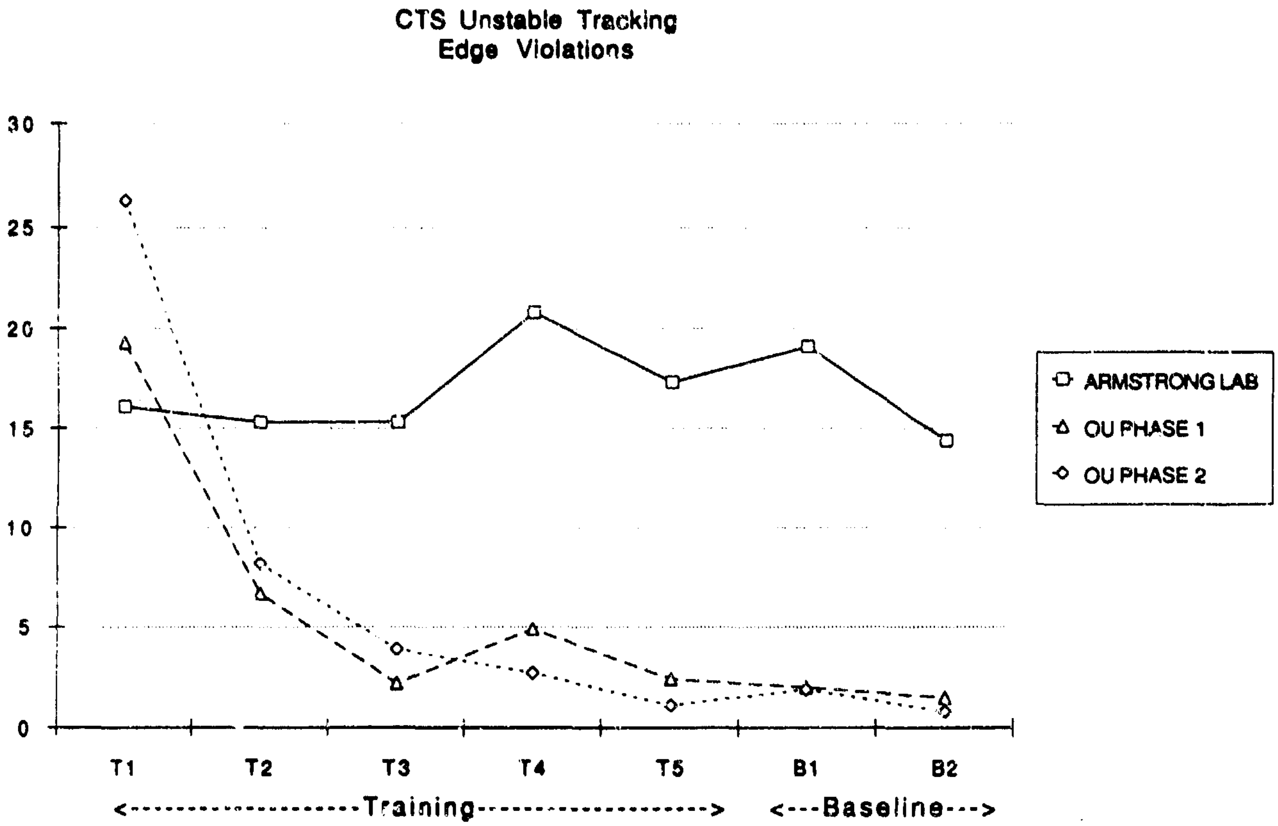

CTS Memory Search-4
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2



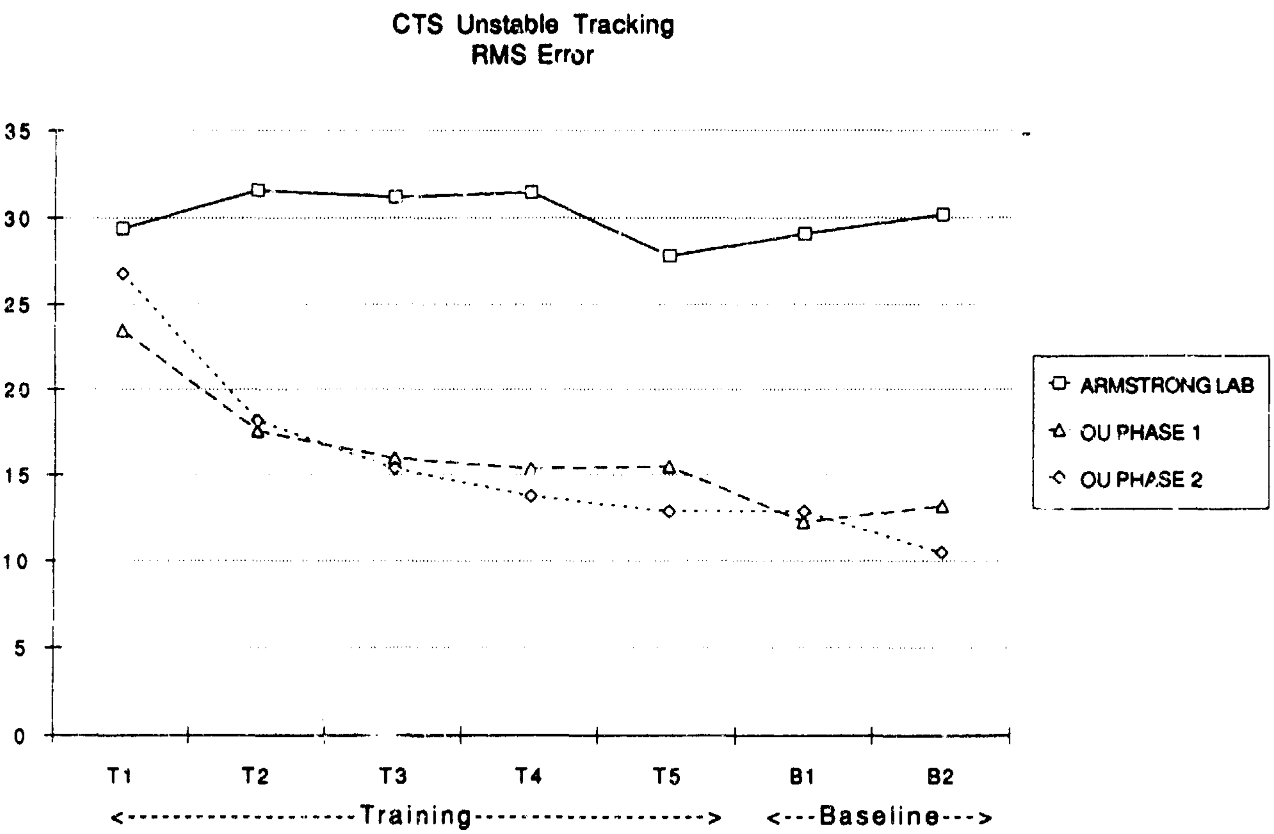
CTS Spatial Processing
Mean Response Time
(msec)
1100
1000
900
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------> <---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2



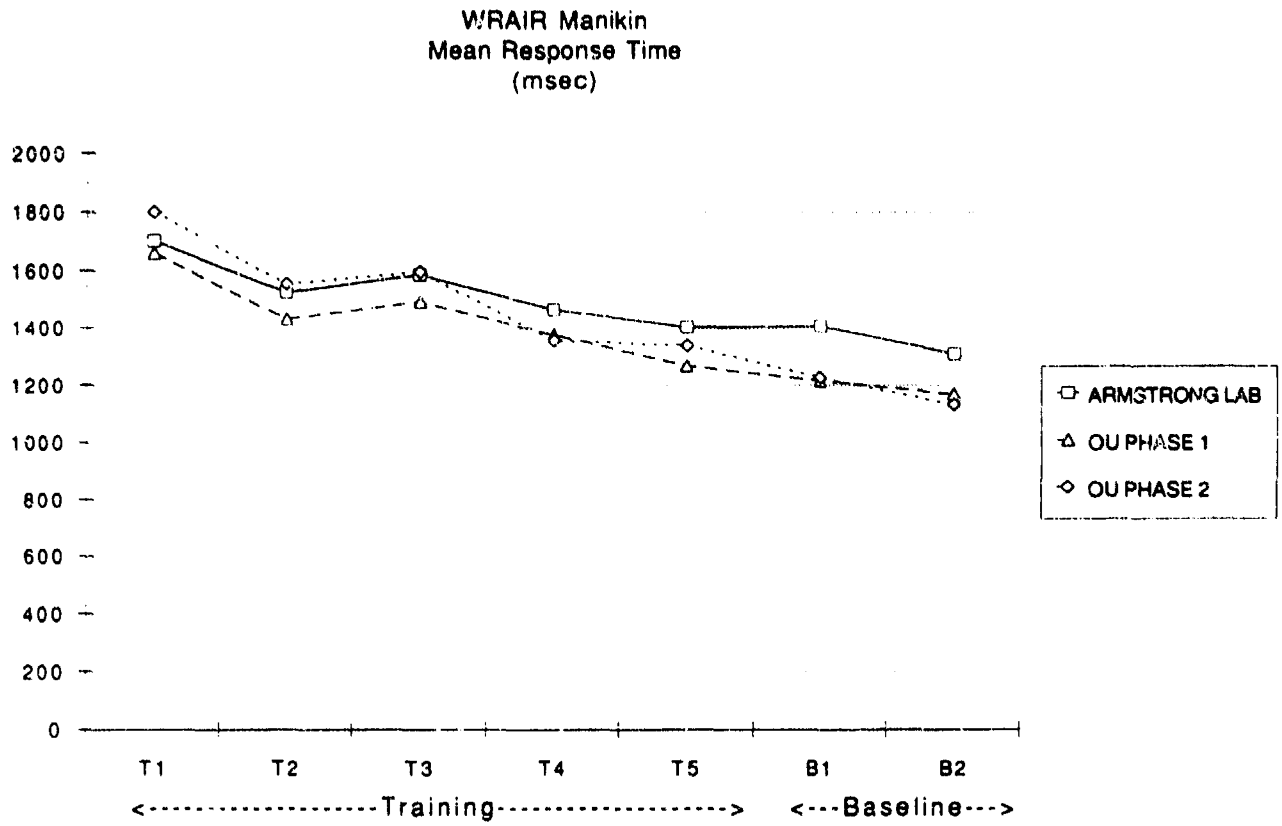
CTS Spatial Processing
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------> <---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2



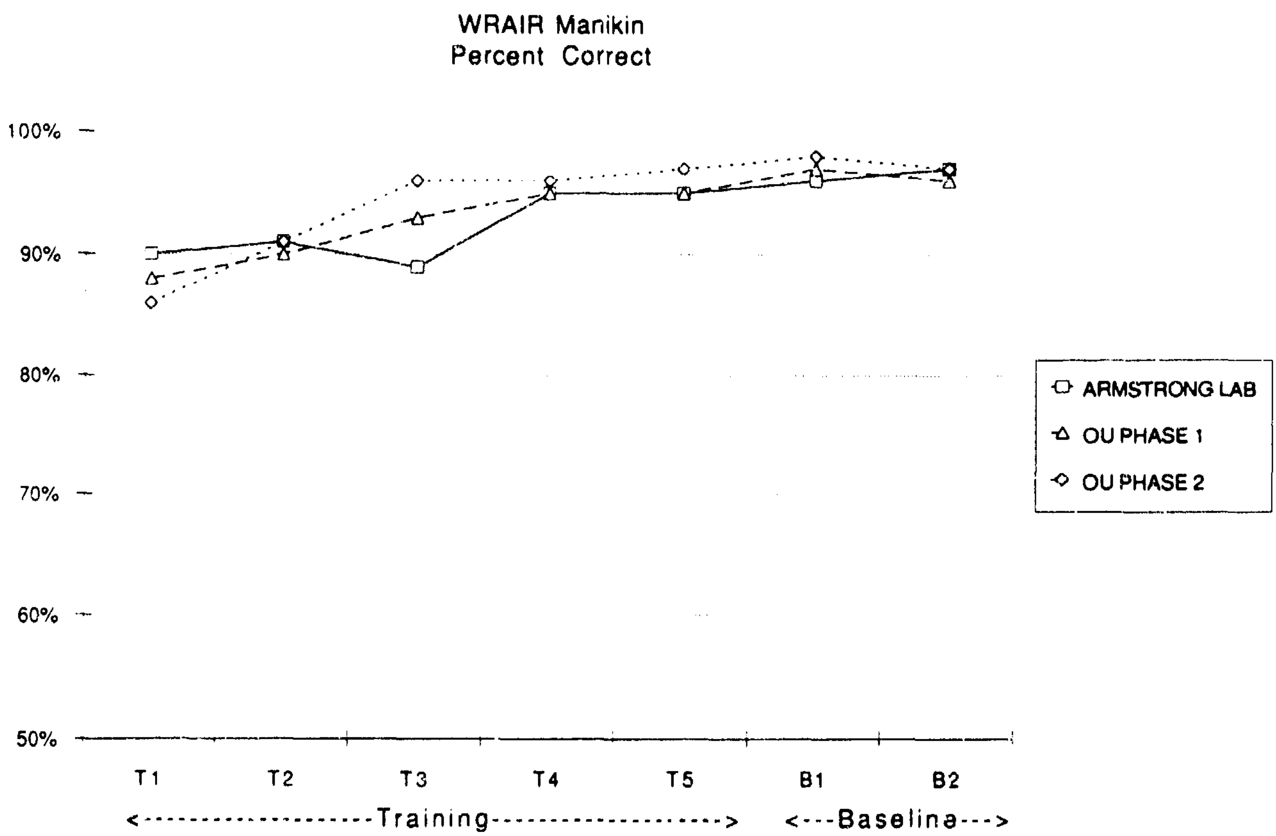
CTS Unstable Tracking
Edge Violations
30
25
20
15
10
5
0
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2



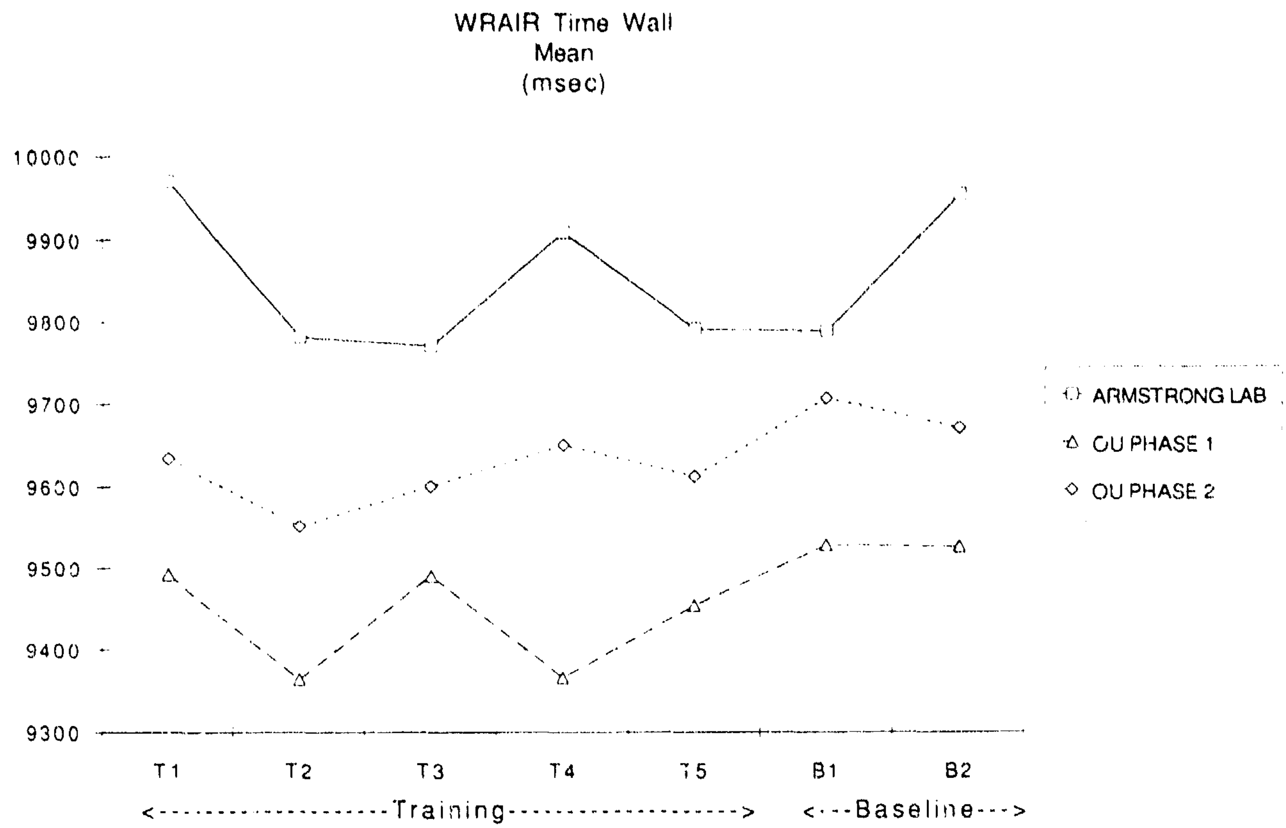
CTS Unstable Tracking
RMS Error
35
30
25
20
15
10
5
0
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

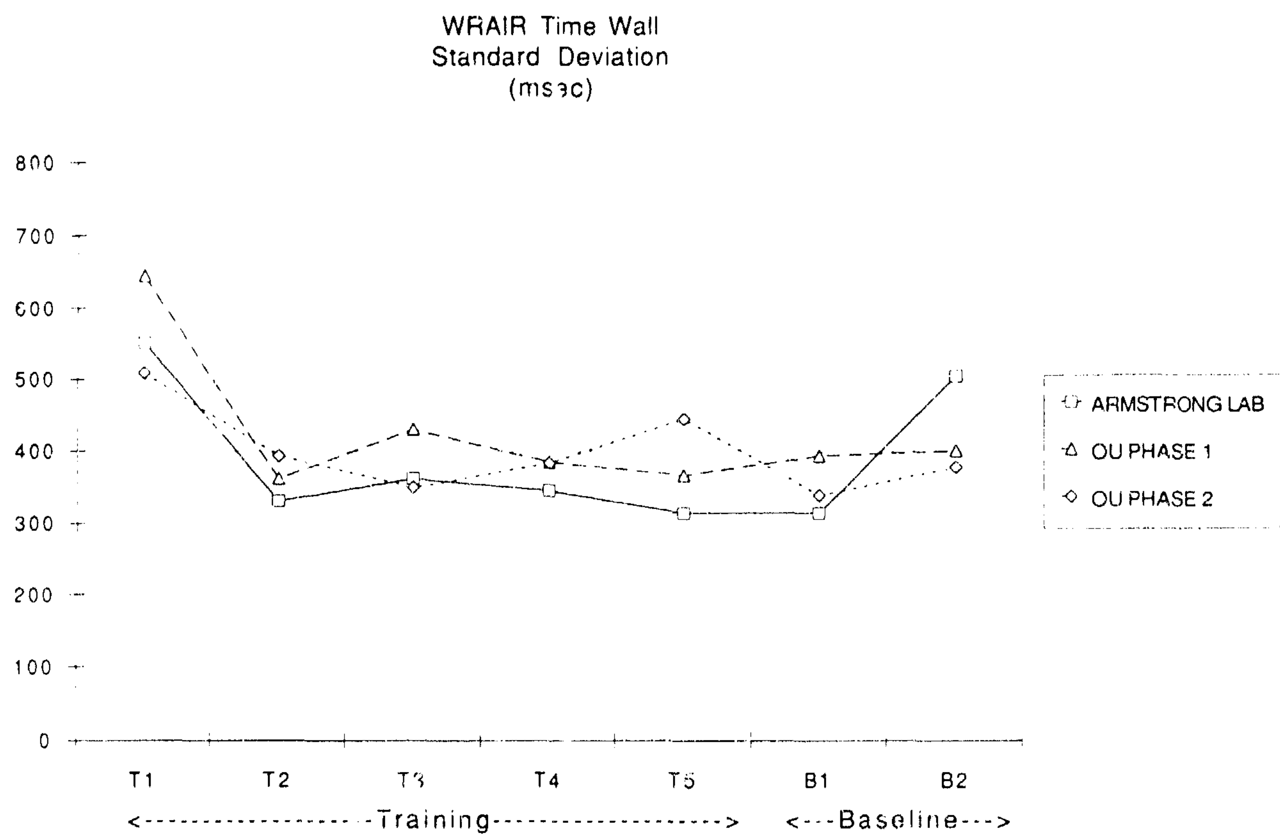


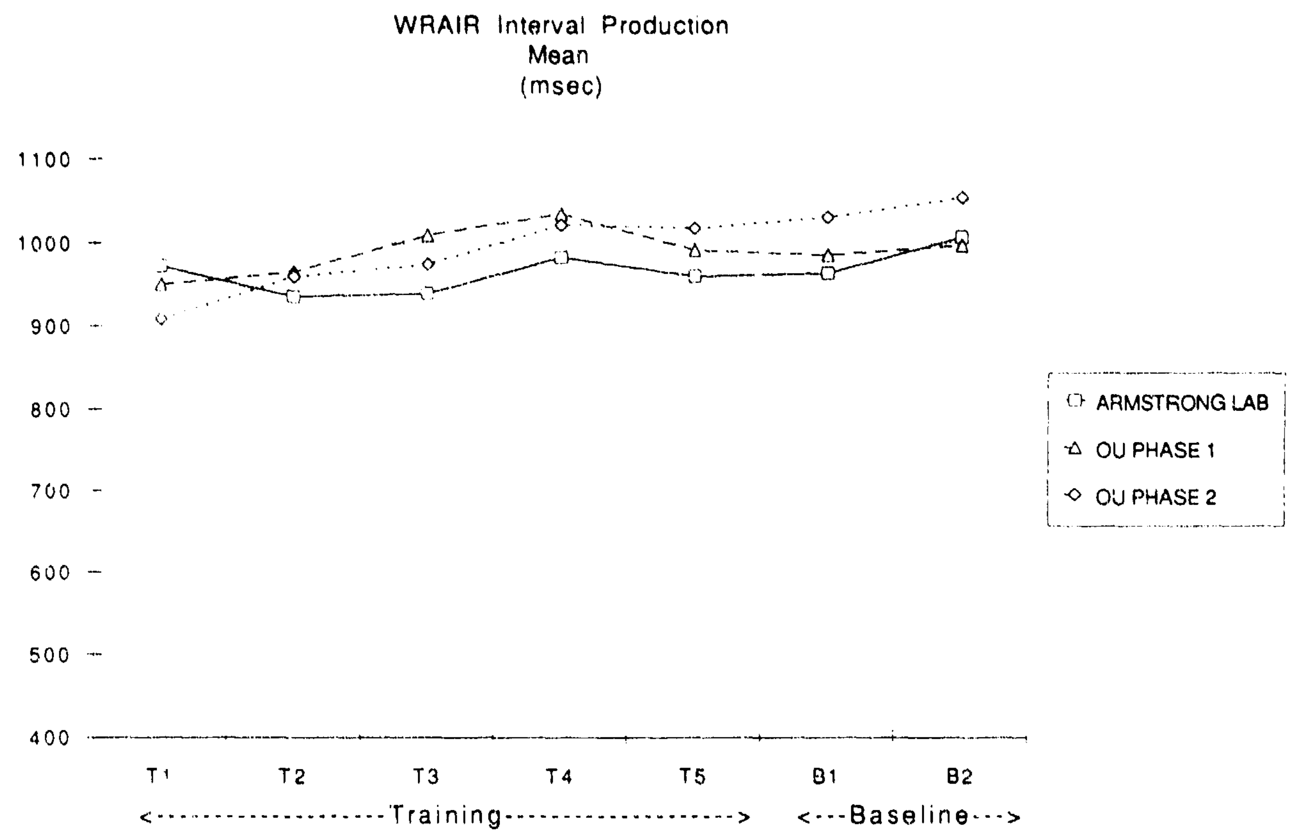
WRAIR Manikin
Mean Response Time
(msec)
2000
1800
1600
1400
1200
1000
800
600
400
200
0
T1
T2
T3
T4
T5
B1
B2
<--------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

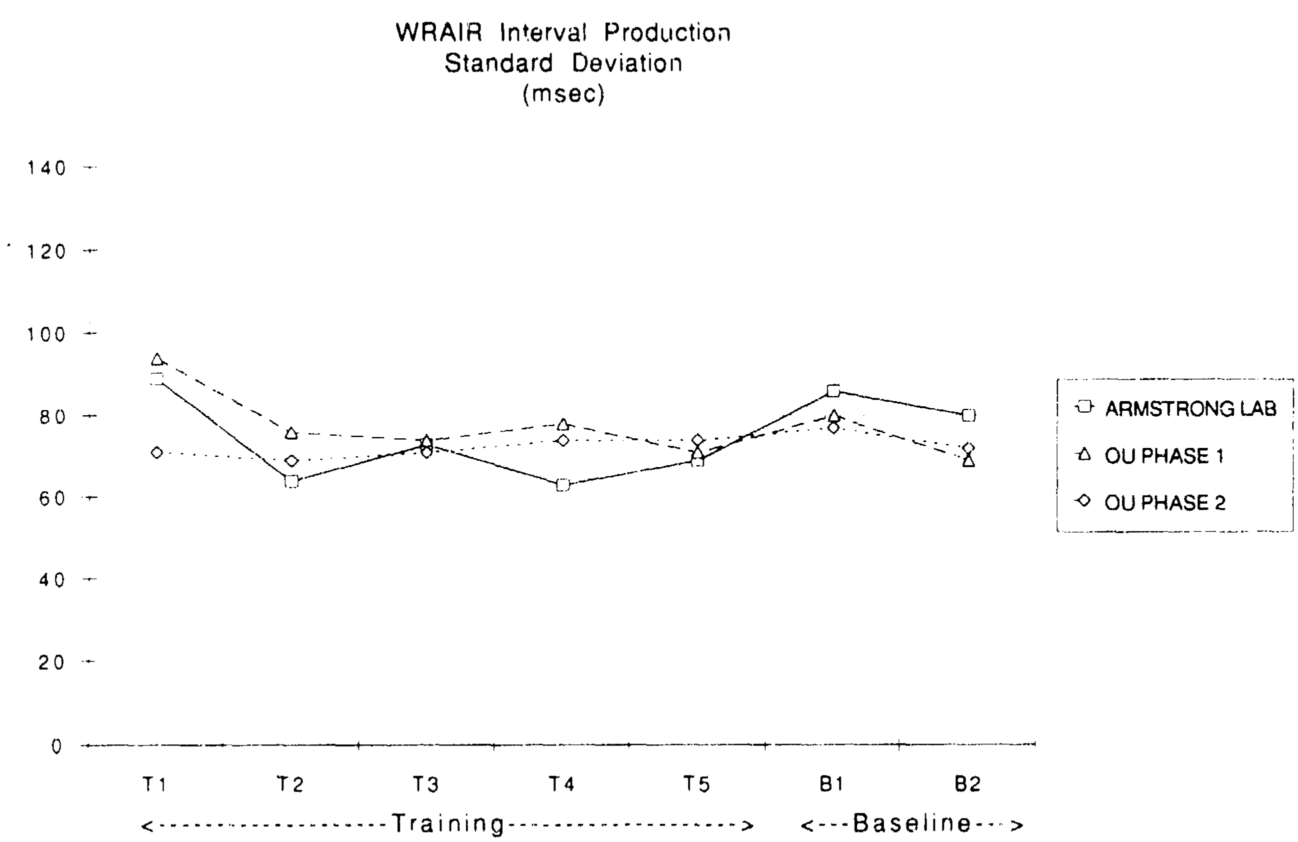


WRAIR Manikin
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2


WRAIR Time Wall
Mean
(msec)
10000
9900
9800
9700
9600
9500
9400
9300
T1
T2
T3
T4
T5
B1
B2
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2


WRAIR Time Wall
Standard Deviation
(msec)
800
700
600
500
400
300
200
100
0
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2


WRAIR Interval Production
Mean
(msec)
1100
1000
900
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

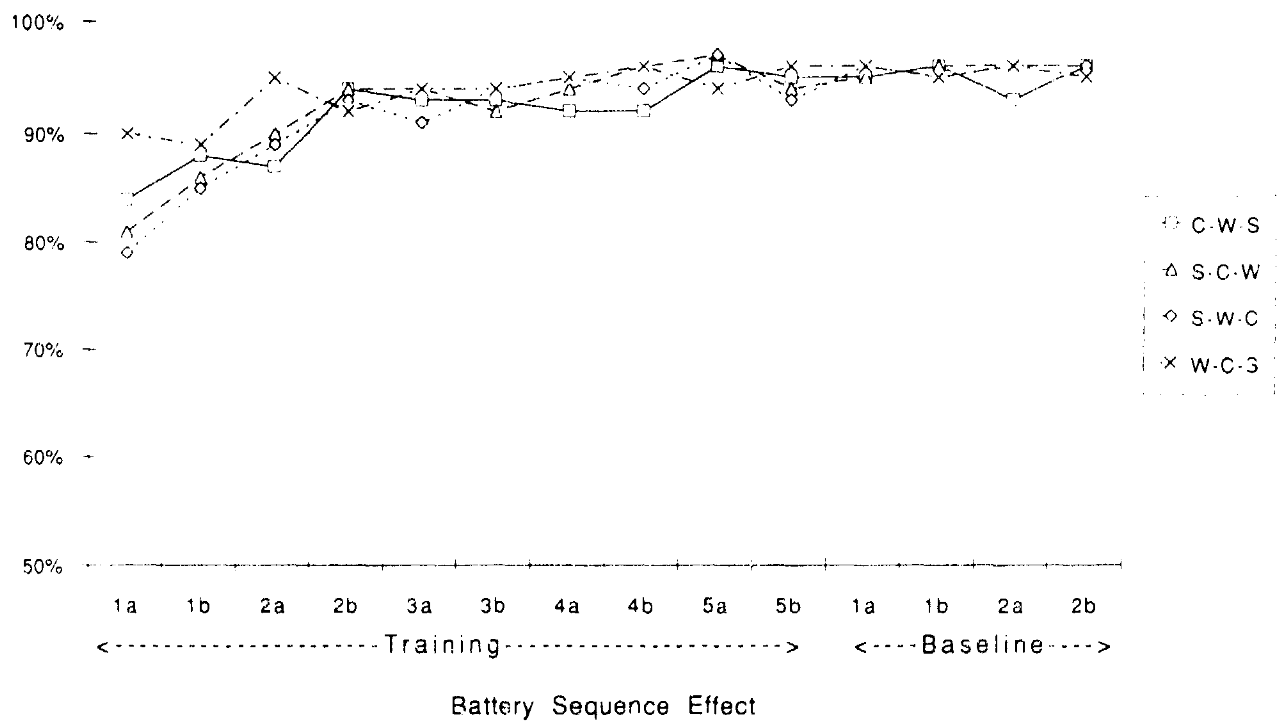

WRAIR Interval Production
Standard Deviation
(msec)
140
120
100
80
60
40
20
0
T1
T2
T3
T4
T5
B1
B2
<-------------------Training------------------->
<---Baseline--->
ARMSTRONG LAB
OU PHASE 1
OU PHASE 2

APPENDIX J

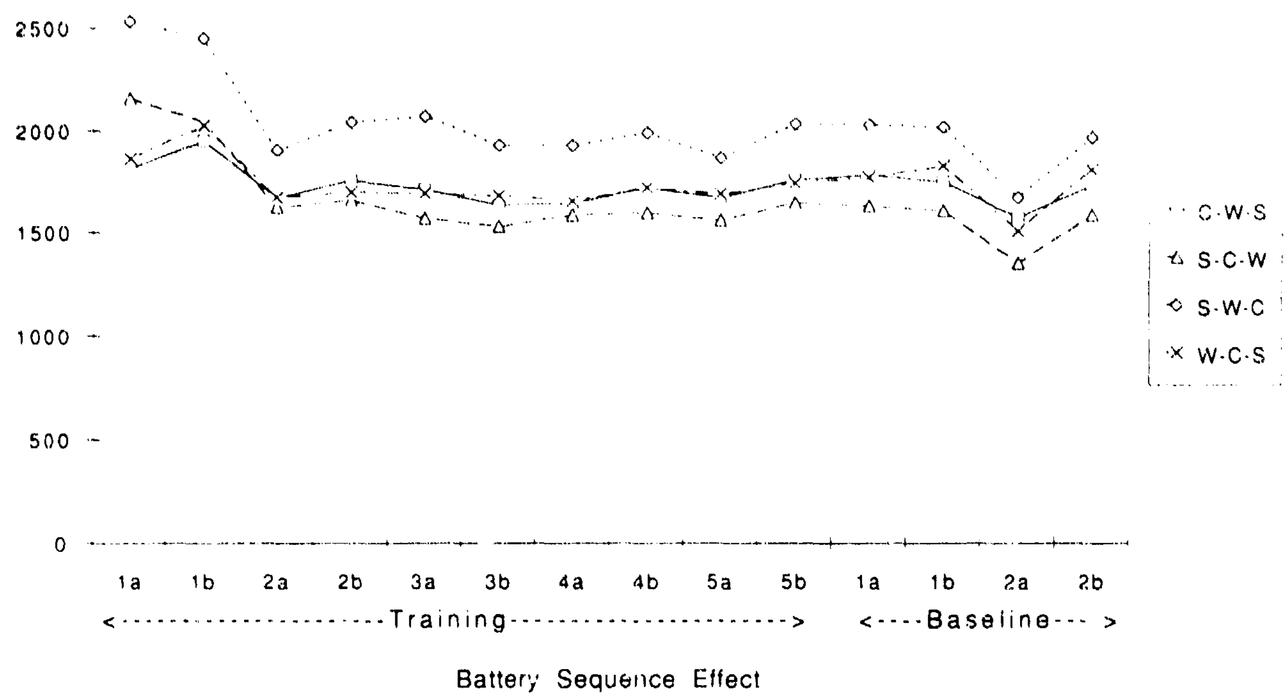
BATTERY SEQUENCE DATA

STRES Grammatical Reasoning
Mean Response Time
(msec)

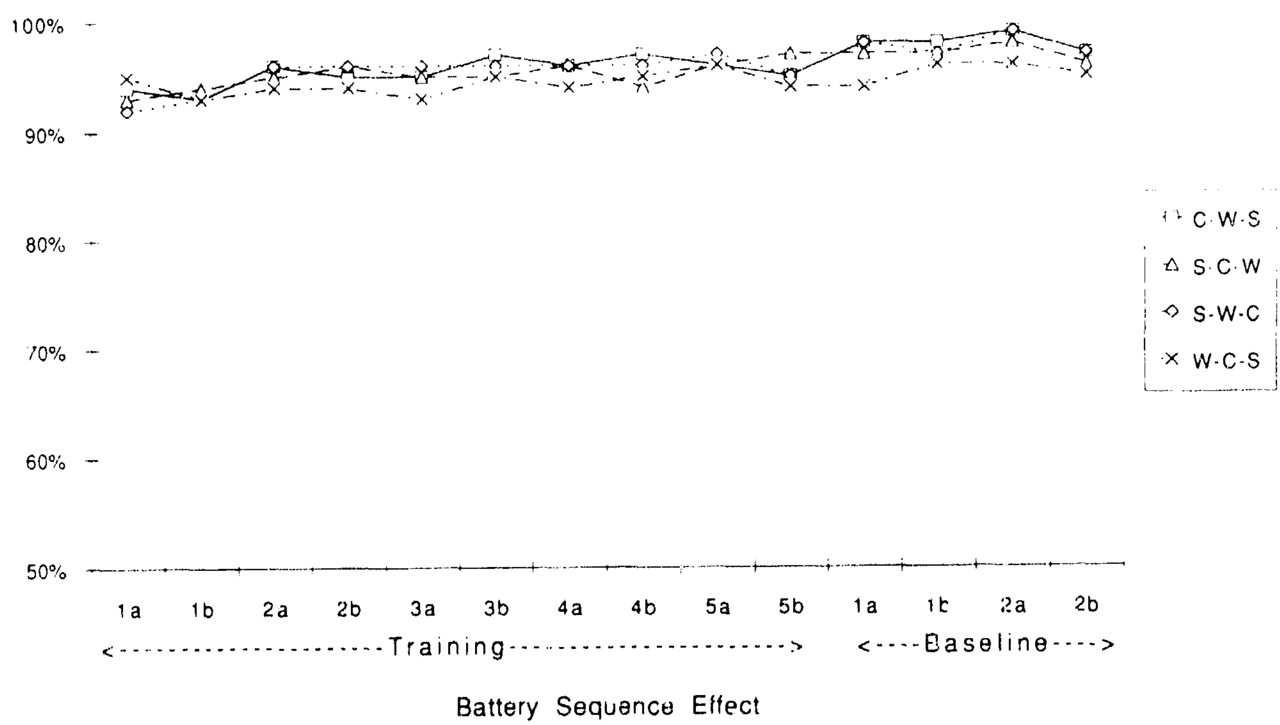
STRES Grammatical Reasoning
Percent Correct

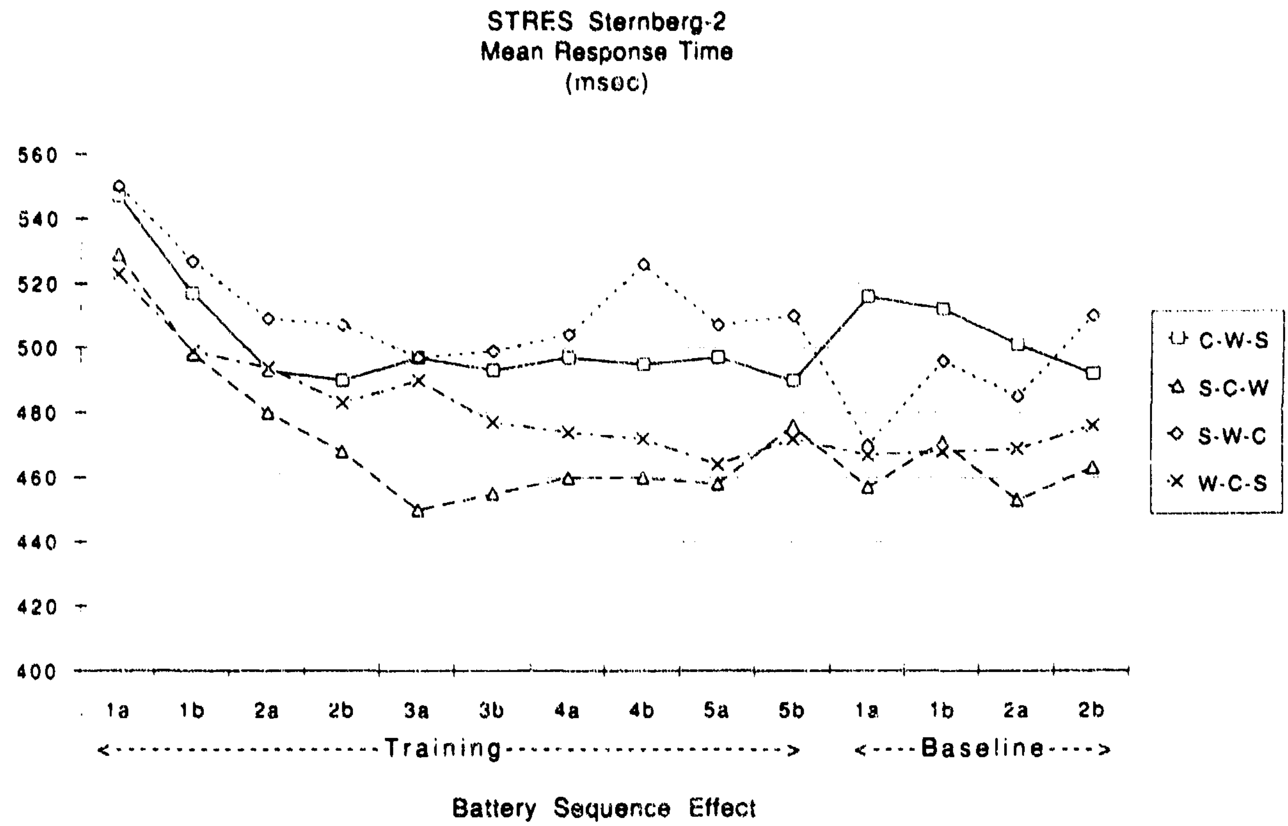


STRES Mathematical Processing
Mean Response Time
(msec)

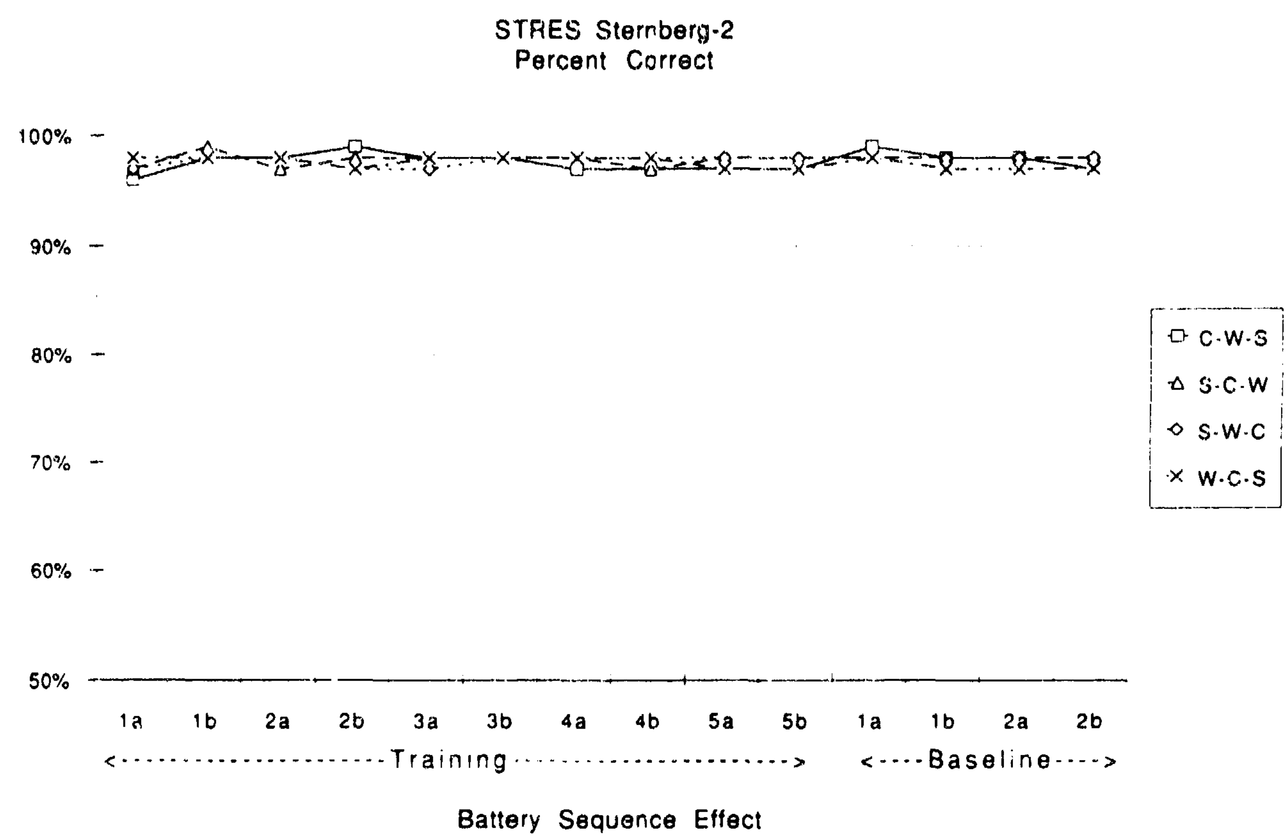


STRES Mathematical Processing
Percent Correct

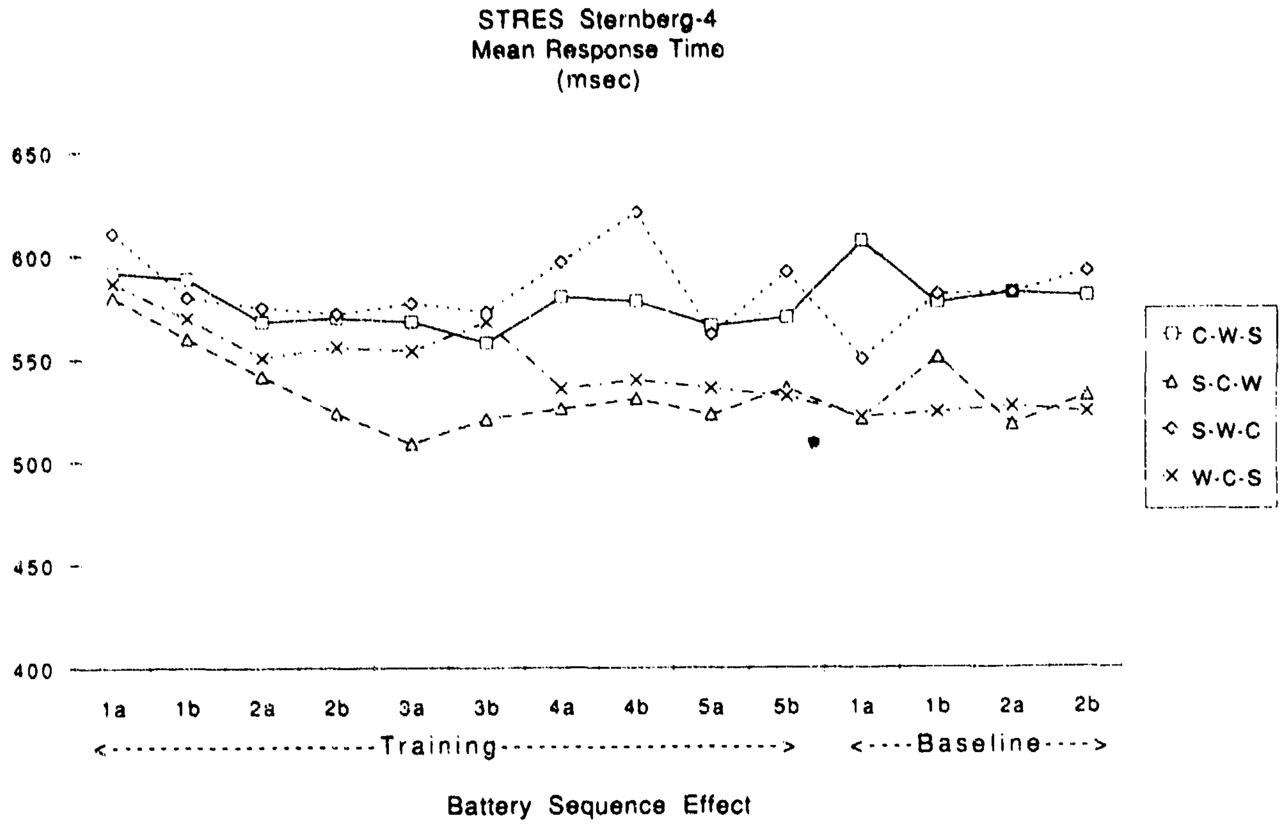


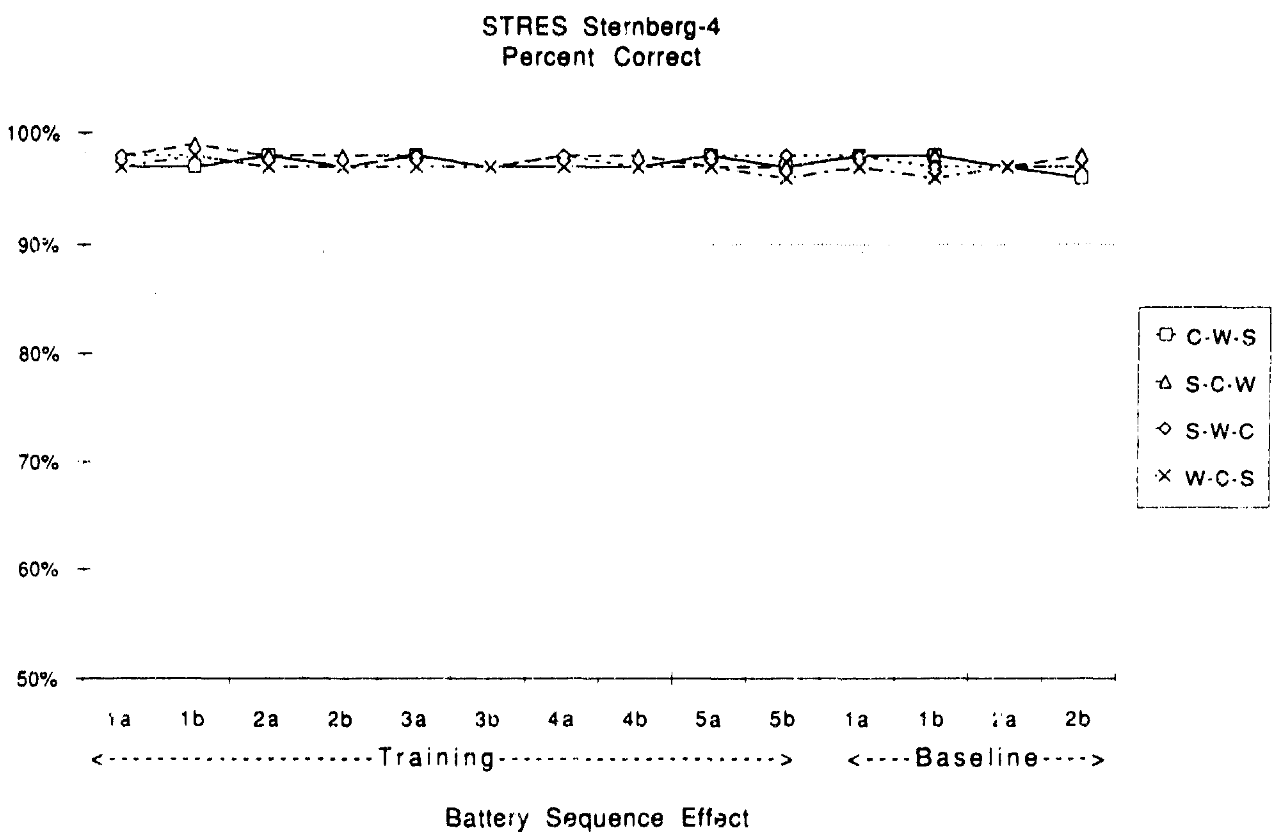


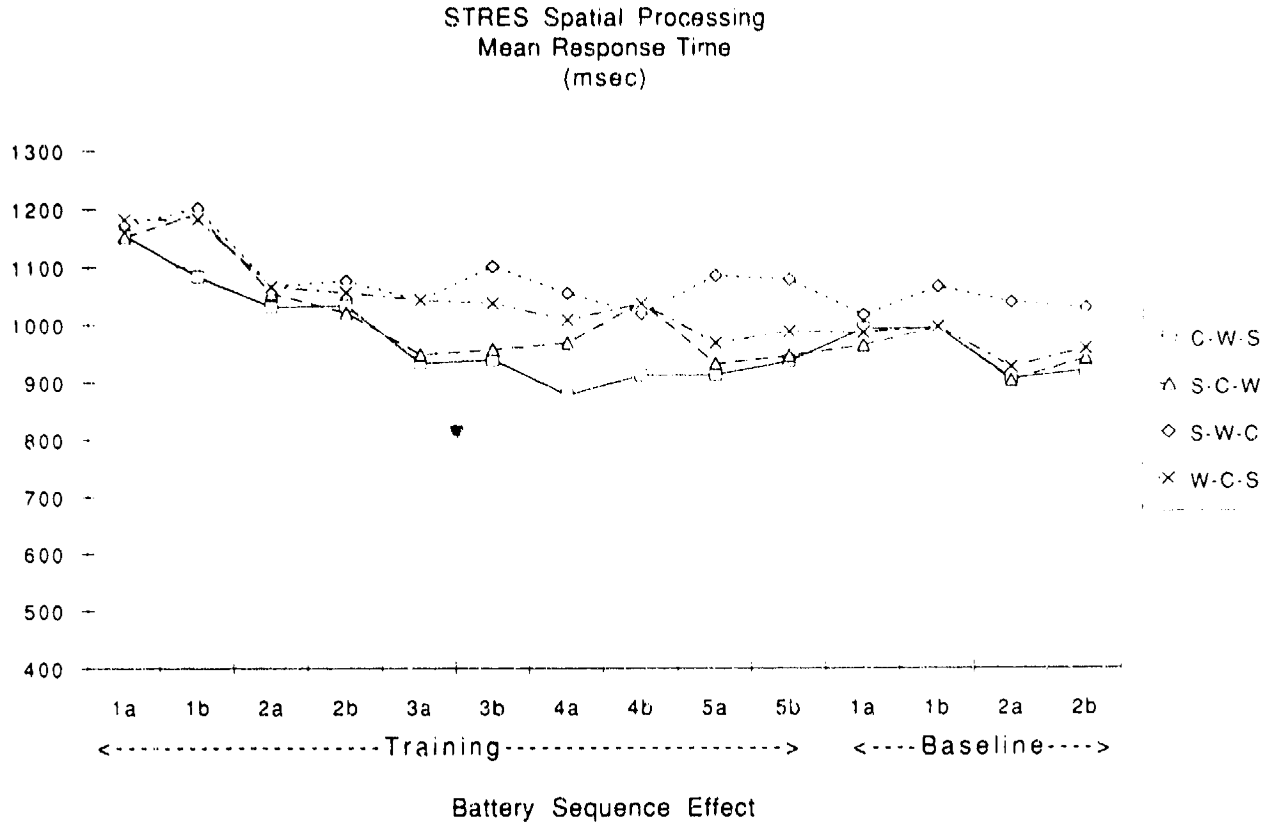
STRES Sternberg-2
Mean Response Time
(msec)
560
540
520
500
480
460
440
420
400
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<--------------------Training---------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

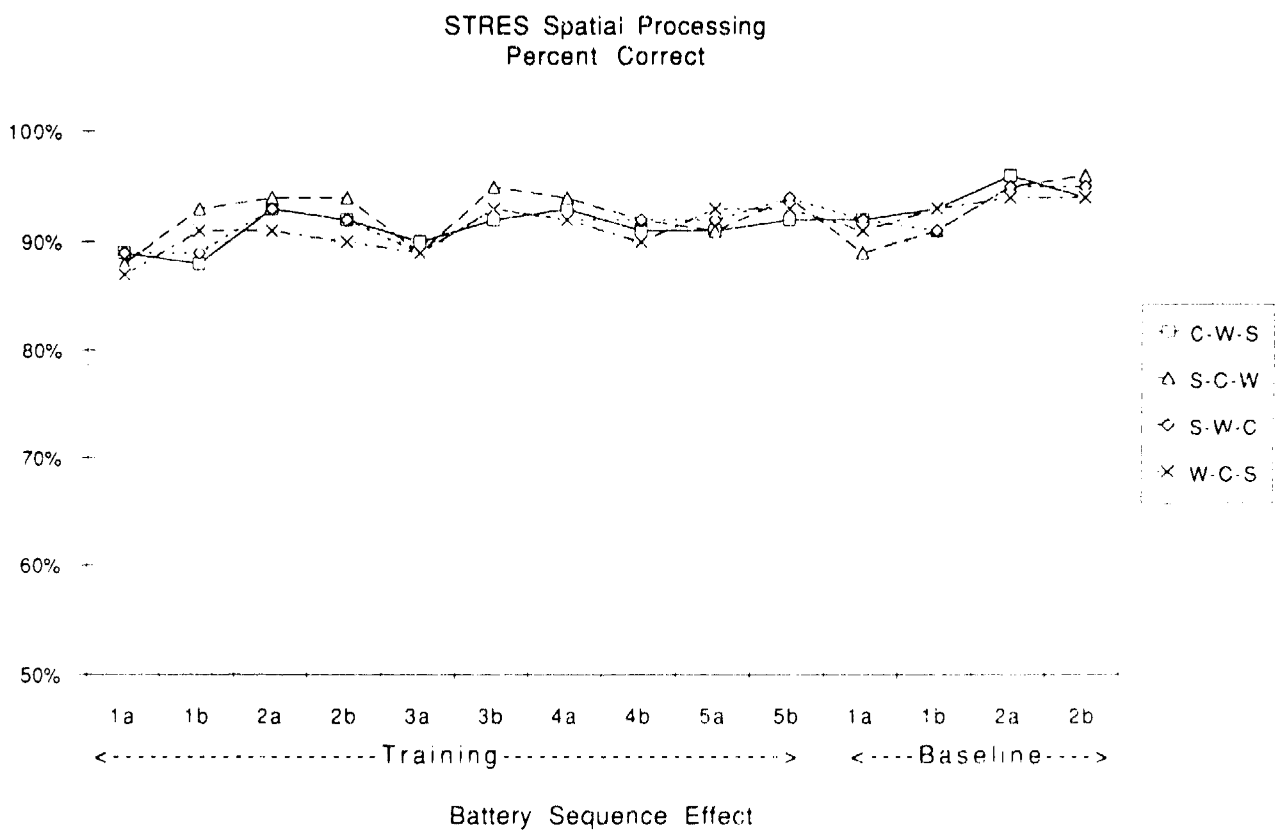


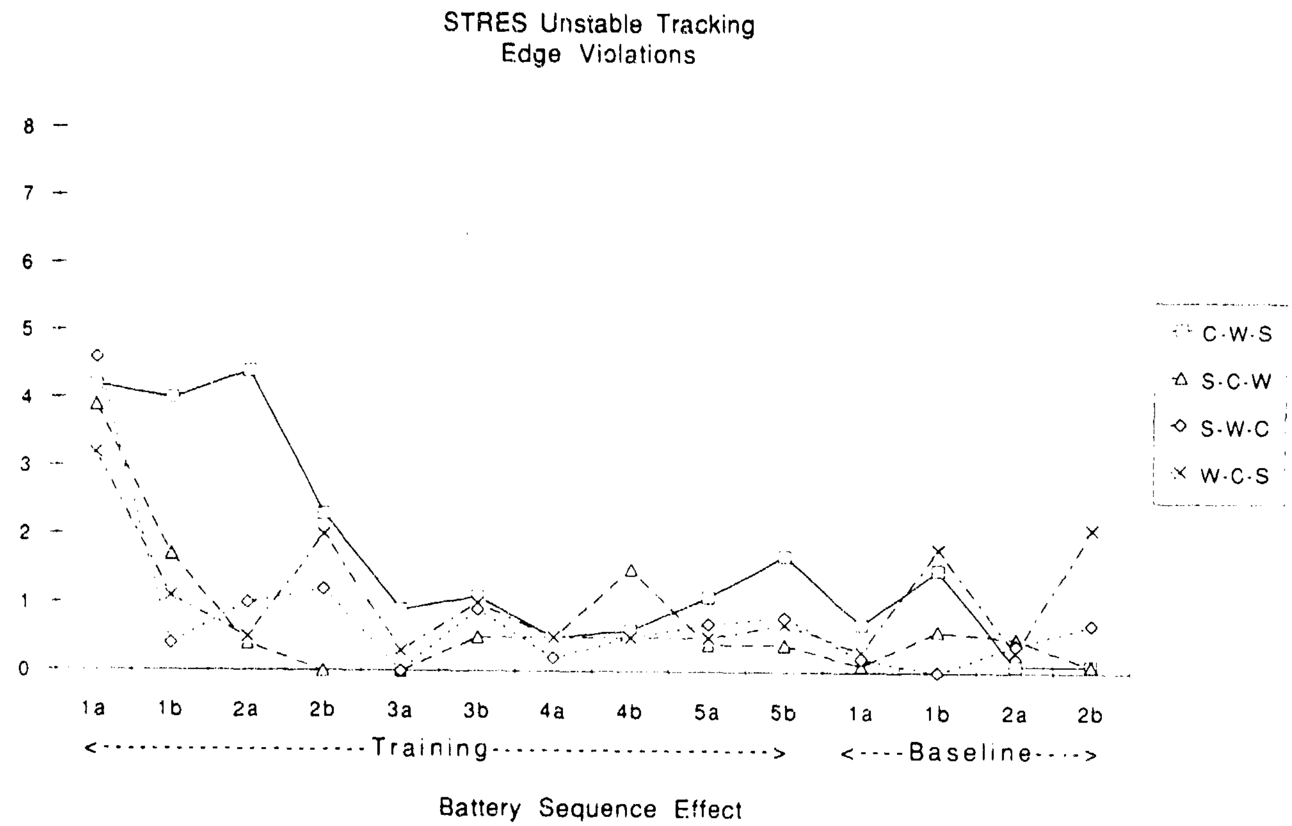
STRES Sternberg-2
Percent Correct
100%
90%
80%
70%
60%
50%
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<--------------------Training---------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S


STRES Sternberg-4
Mean Response Time
(msec)
650
600
550
500
450
400
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<---------------------Training---------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

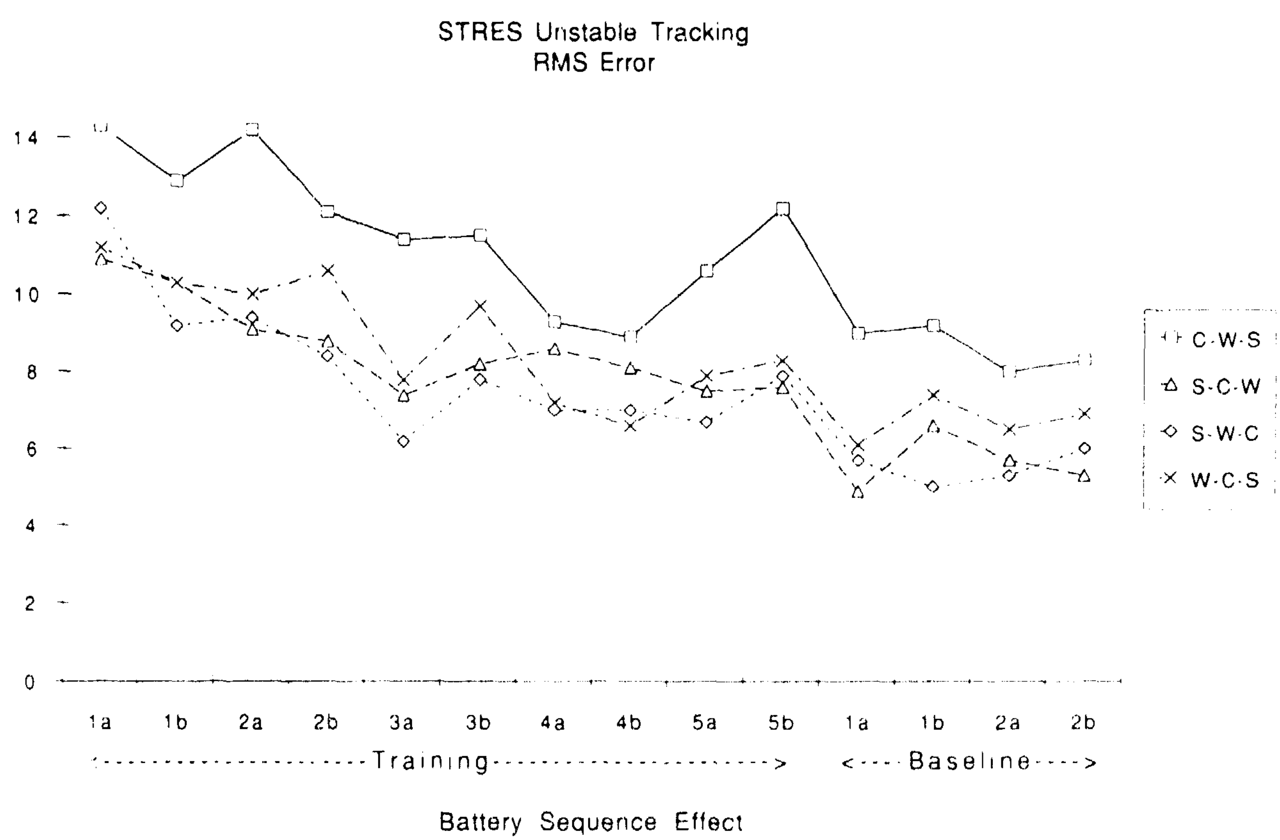

STRES Sternberg-4
Percent Correct
100%
90%
80%
70%
60%
50%
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<---------------------Training---------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S


STRES Spatial Processing
Mean Response Time
(msec)
1300
1200
1100
1000
900
800
700
600
500
400
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<---------------------Training--------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

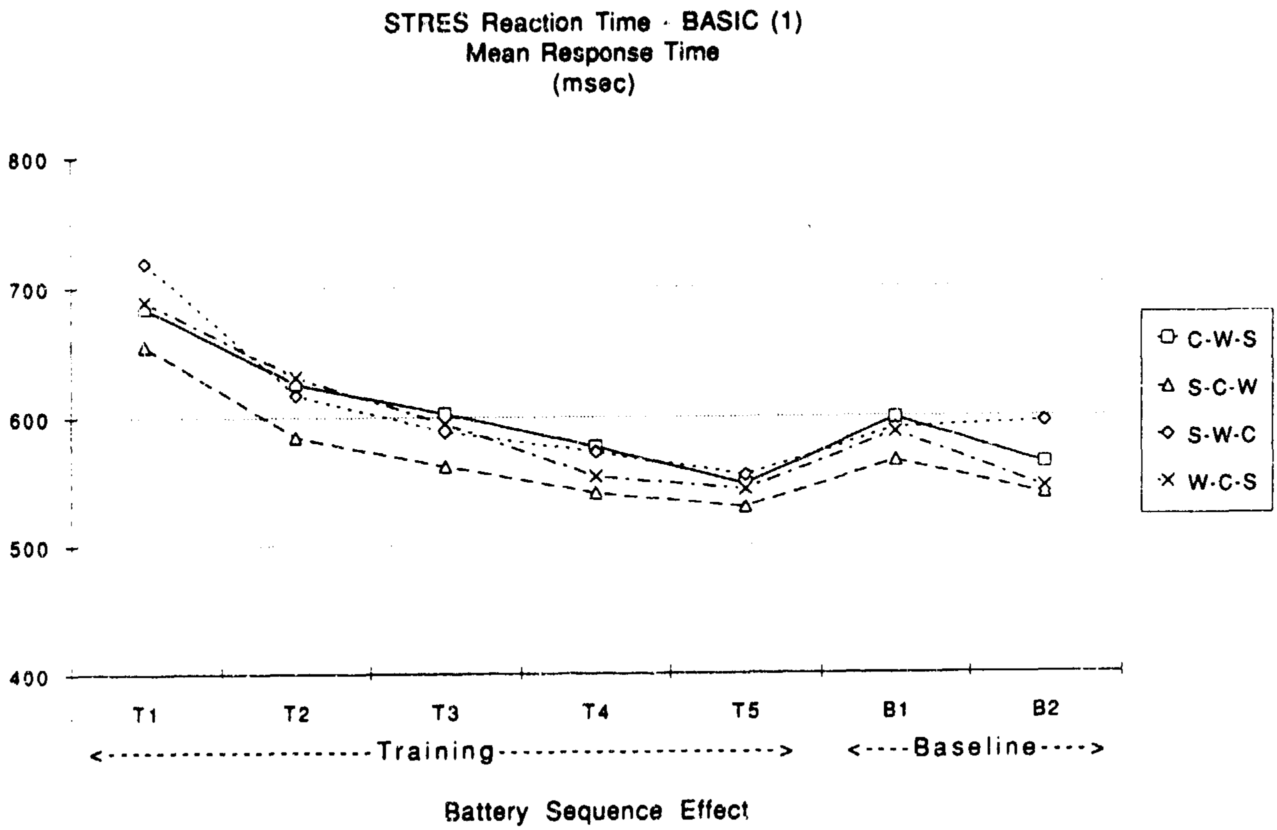

STRES Spatial Processing
Percent Correct
100%
90%
80%
70%
60%
50%
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<---------------------Training--------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

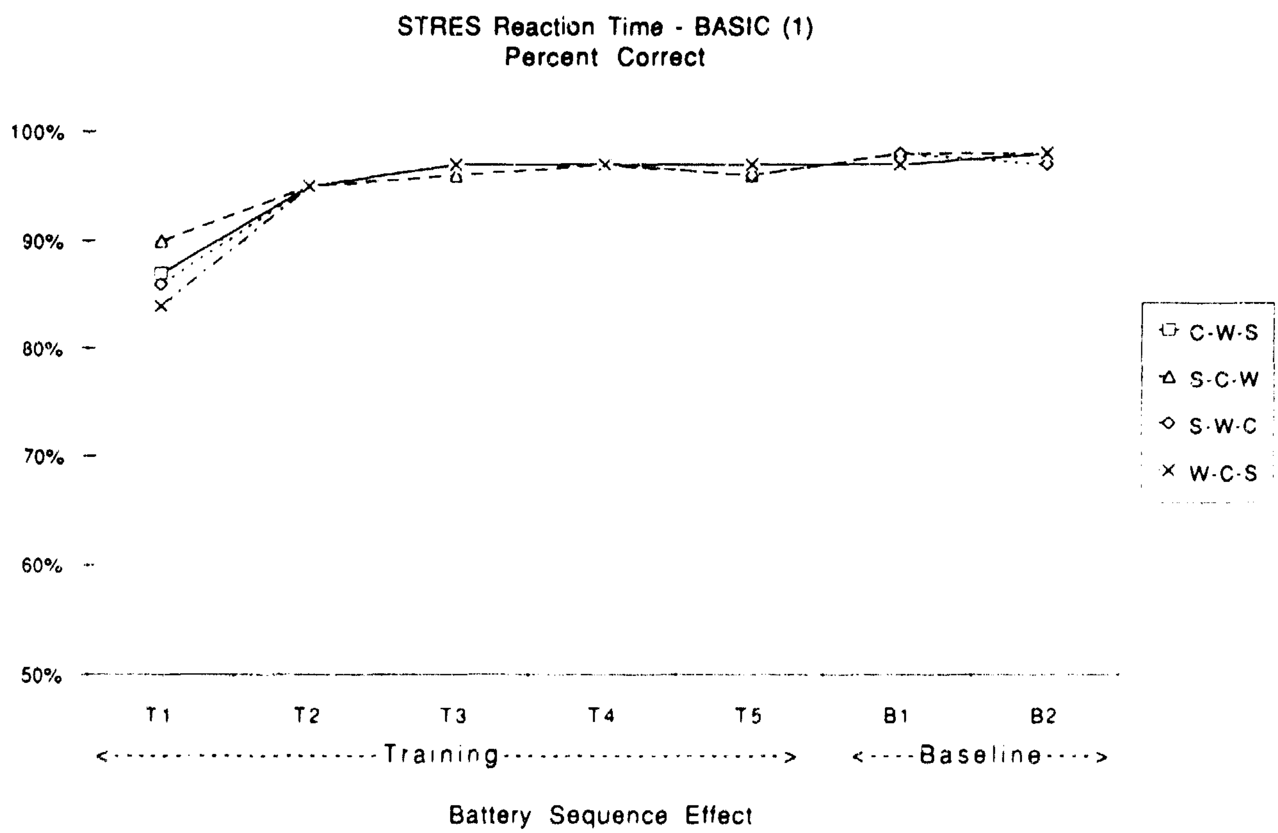


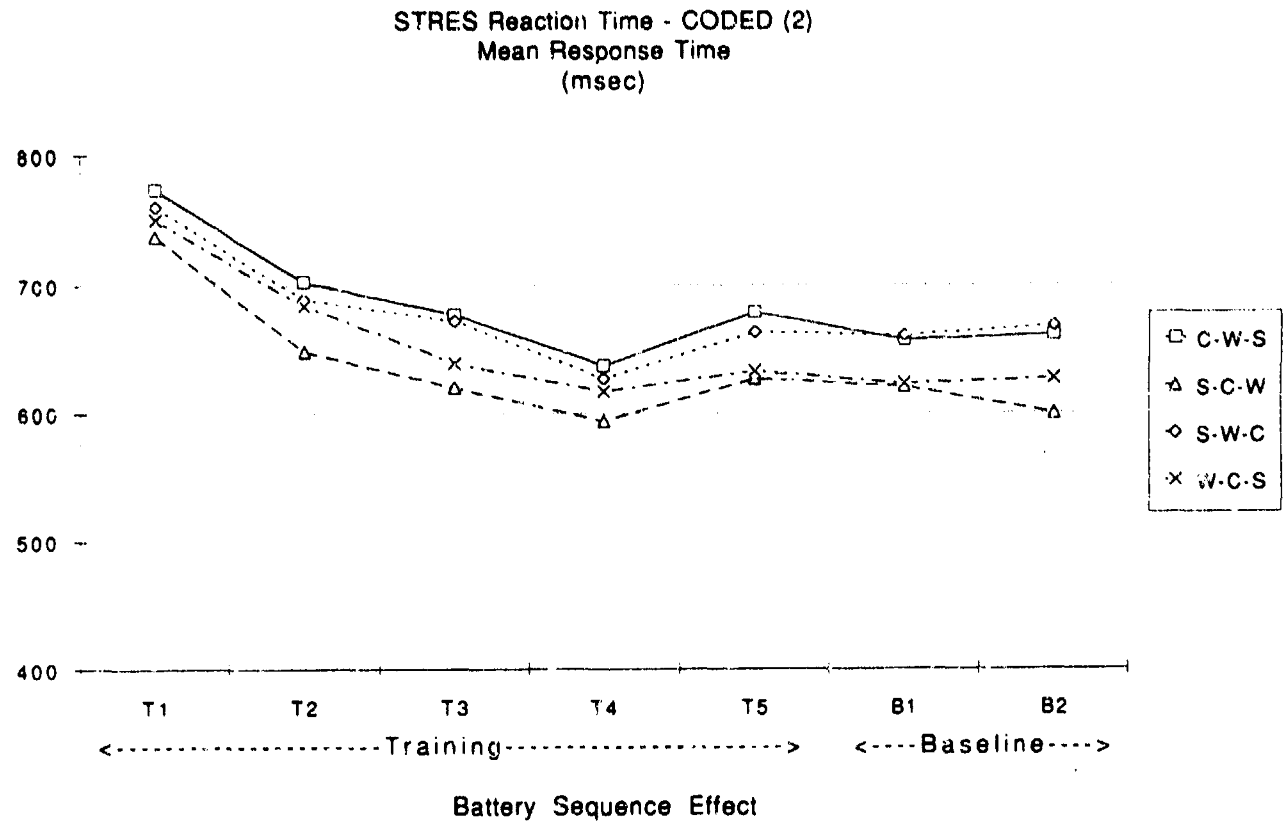
STRES Unstable Tracking
Edge Violations
8
7
6
5
4
3
2
1
0
C-W-S
S-C-W
S-W-C
W-C-S
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<--------------------Training---------------------> <----Baseline---->
Battery Sequence Effect

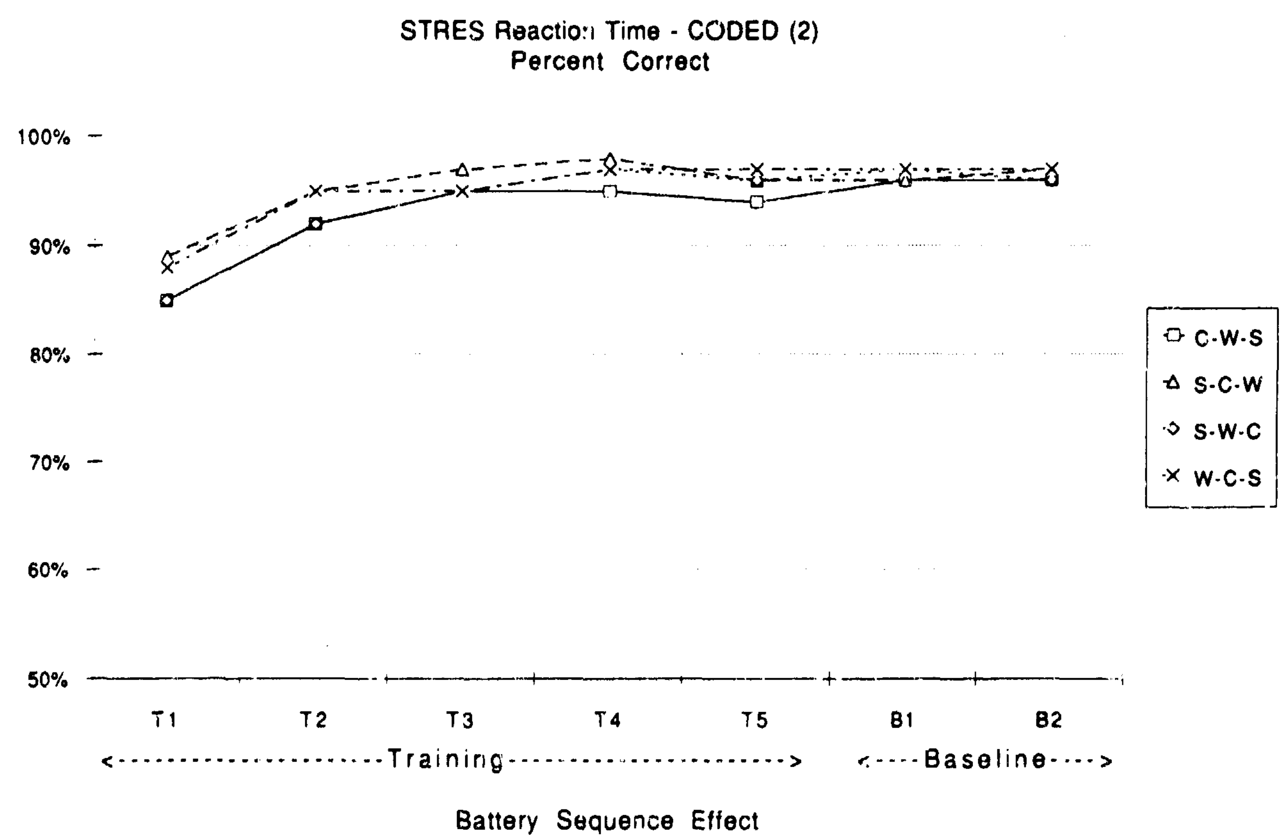


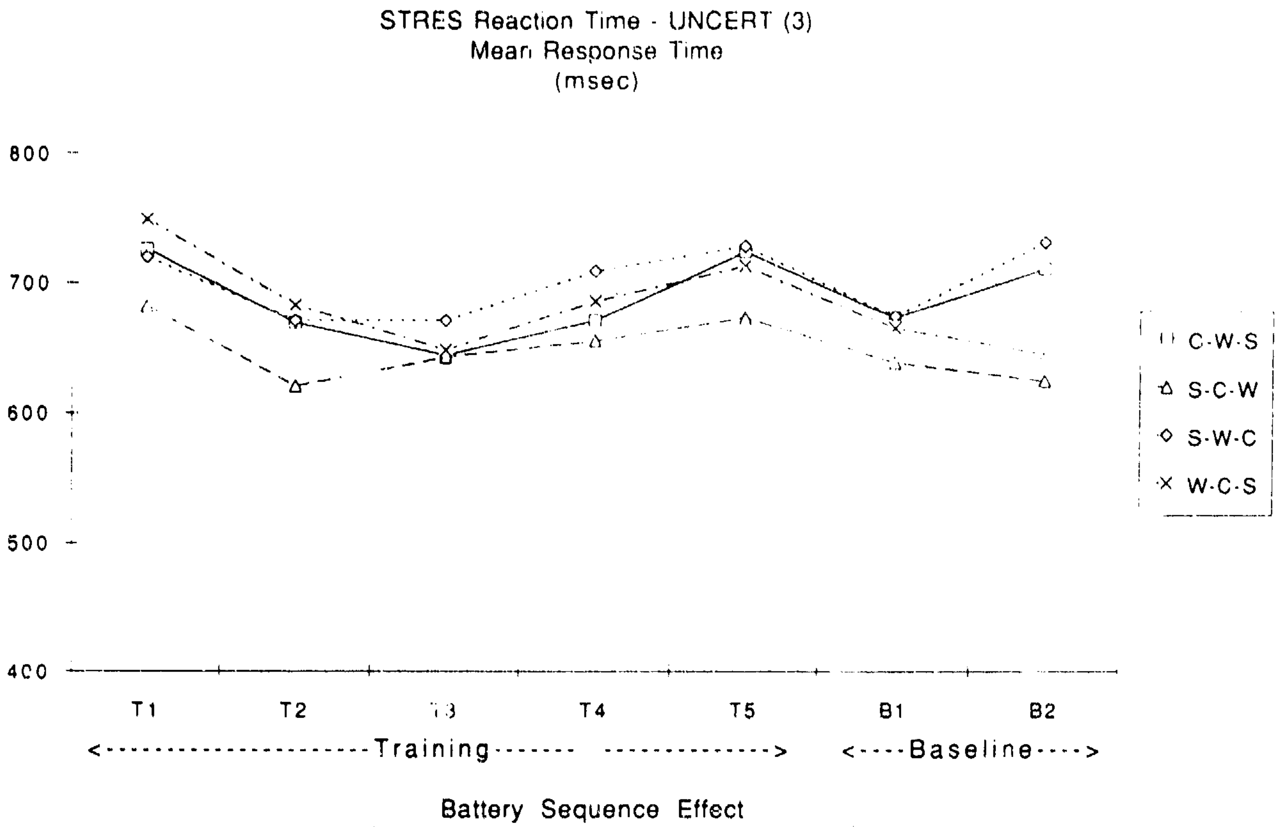
STRES Unstable Tracking
RMS Error
14
12
10
8
6
4
2
0
C-W-S
S-C-W
S-W-C
W-C-S
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<--------------------Training---------------------> <----Baseline---->
Battery Sequence Effect

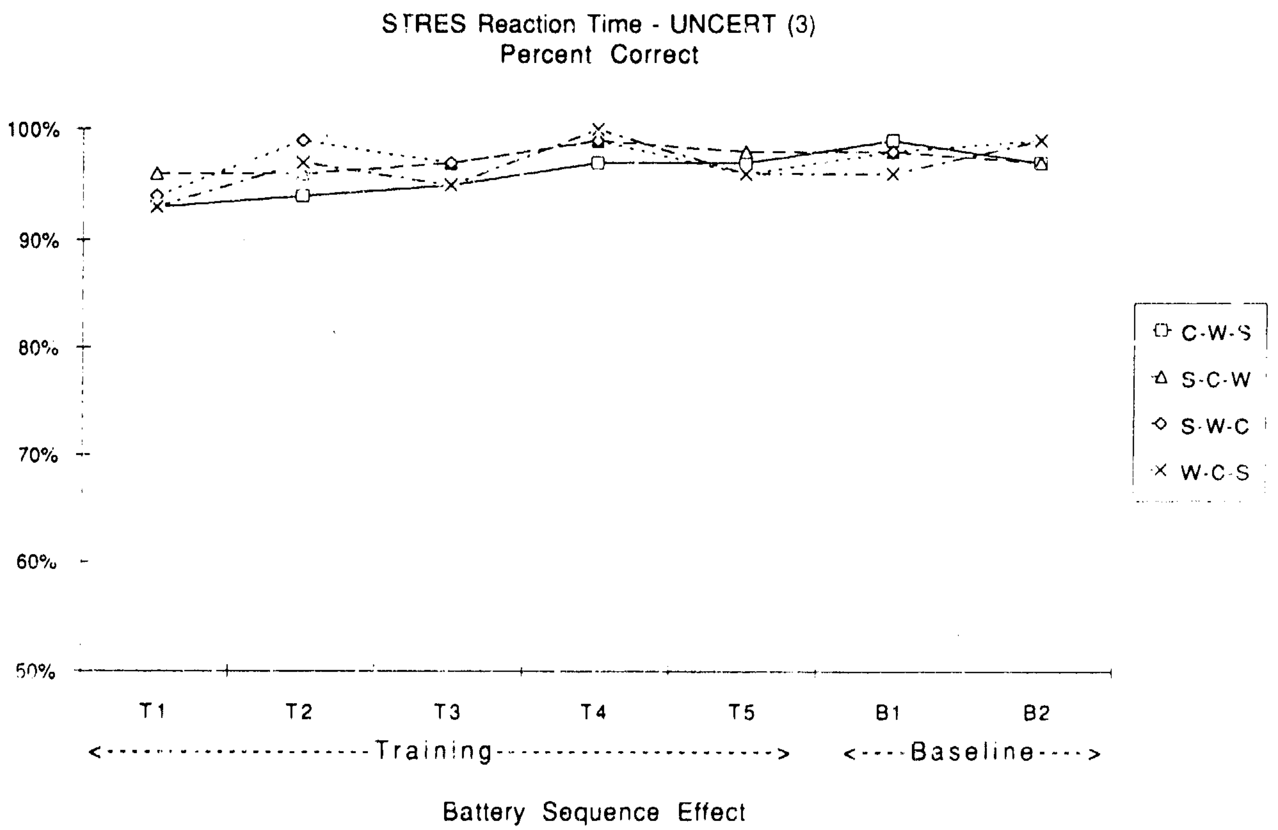

STRES Reaction Time - BASIC (1)
Mean Response Time
(msec)
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<--------------------Training-------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

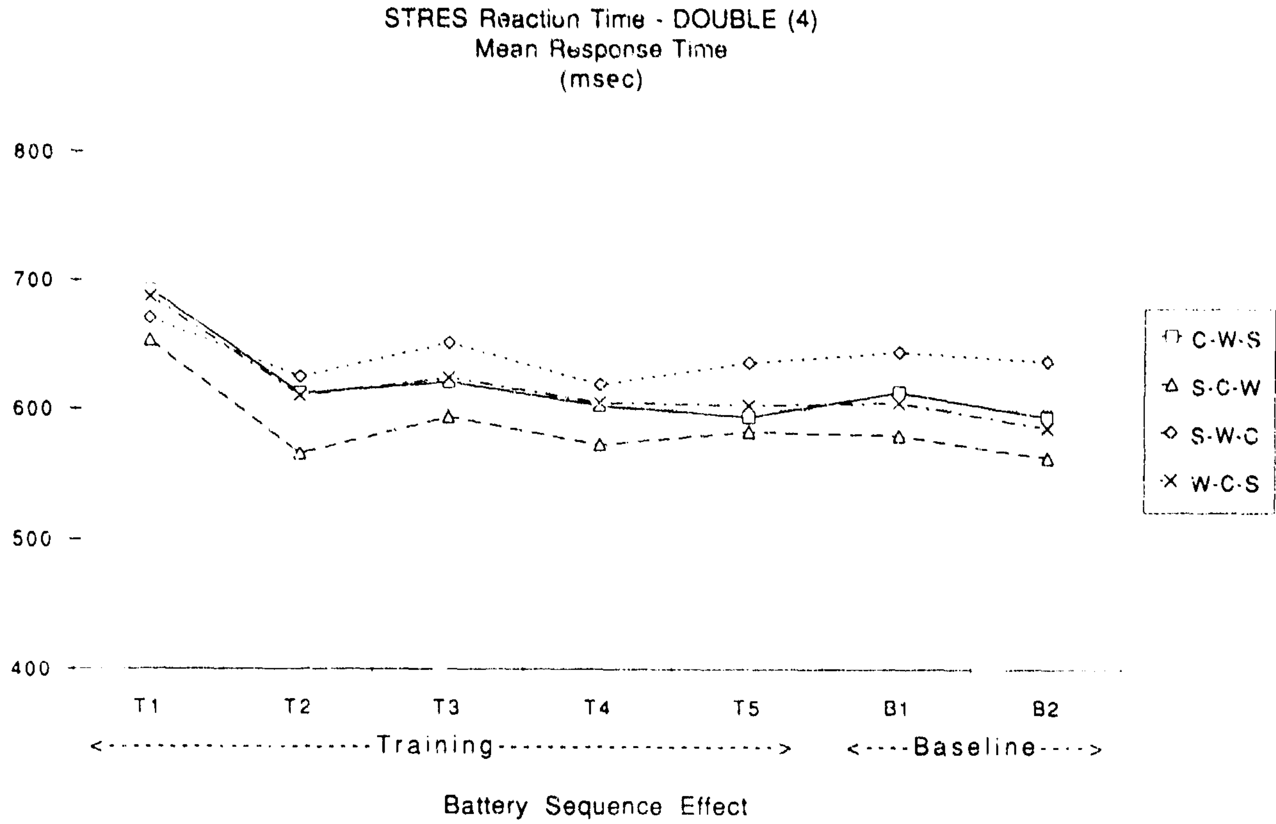

STRES Reaction Time - BASIC (1)
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<--------------------Training-------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

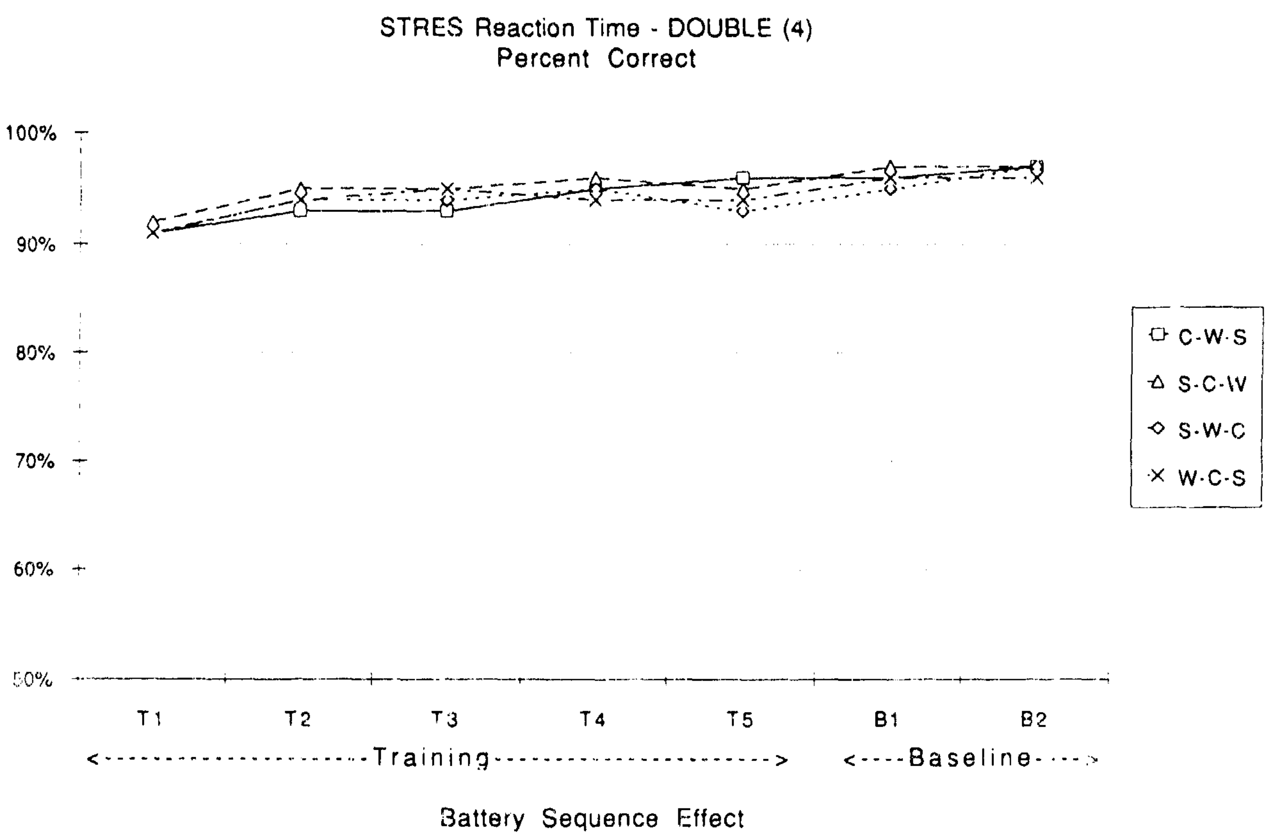

STRES Reaction Time - CODED (2)
Mean Response Time
(msec)
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<-----Training----->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S


STRES Reaction Time - CODED (2)
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<-----Training----->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

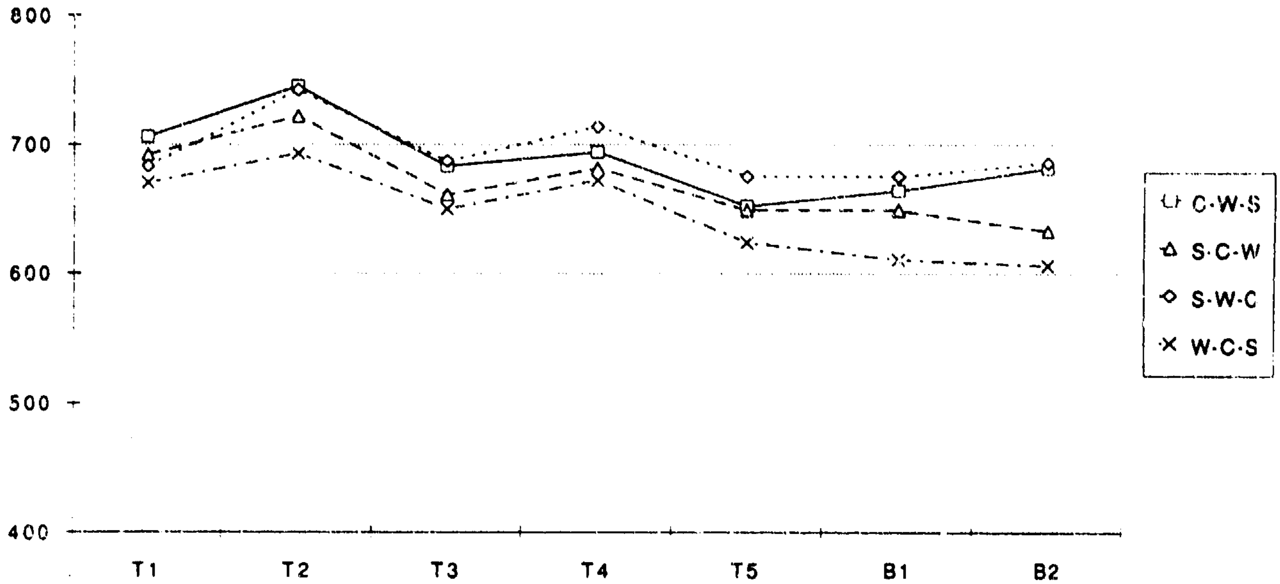

STRES Reaction Time - UNCERT (3)
Mean Response Time
(msec)
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<---------------------Training-------------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S


STRES Reaction Time - UNCERT (3)
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<---------------------Training---------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S


STRES Reaction Time - DOUBLE (4)
Mean Response Time
(msec)
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<--------------------Training---------------------> <----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S


STRES Reaction Time - DOUBLE (4)
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<--------------------Training---------------------> <----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

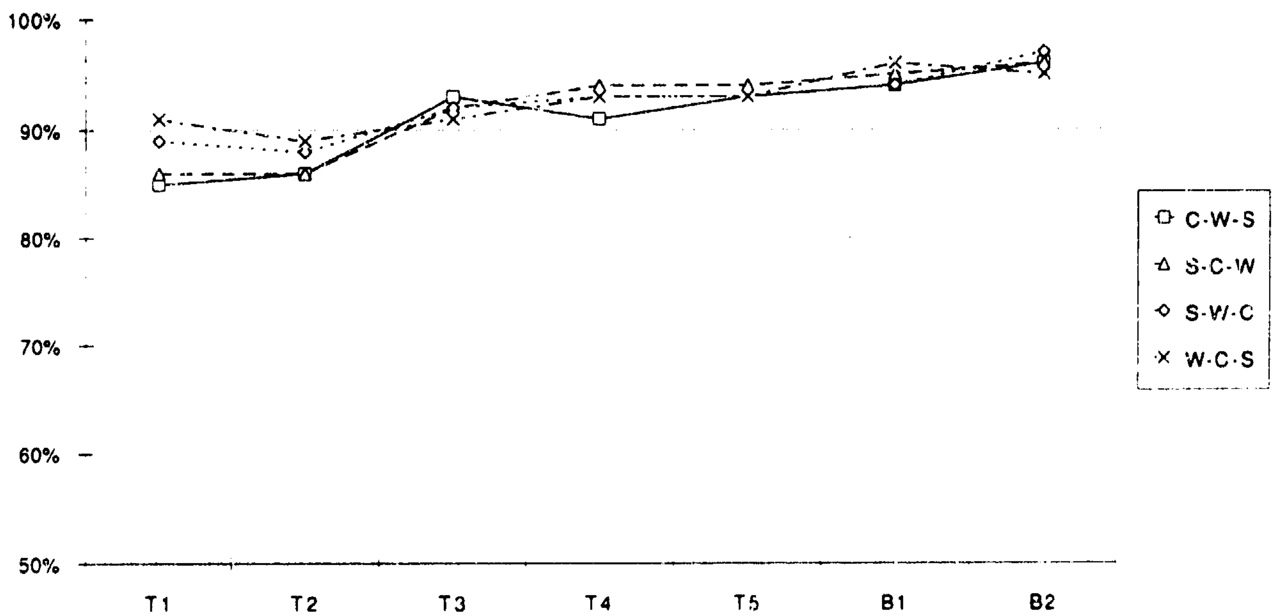
STRES Reaction Time - INVERT (5)
Mean Response Time
(msec)



<-------------------Training-------------------> <----Baseline---->

Battery Sequence Effect

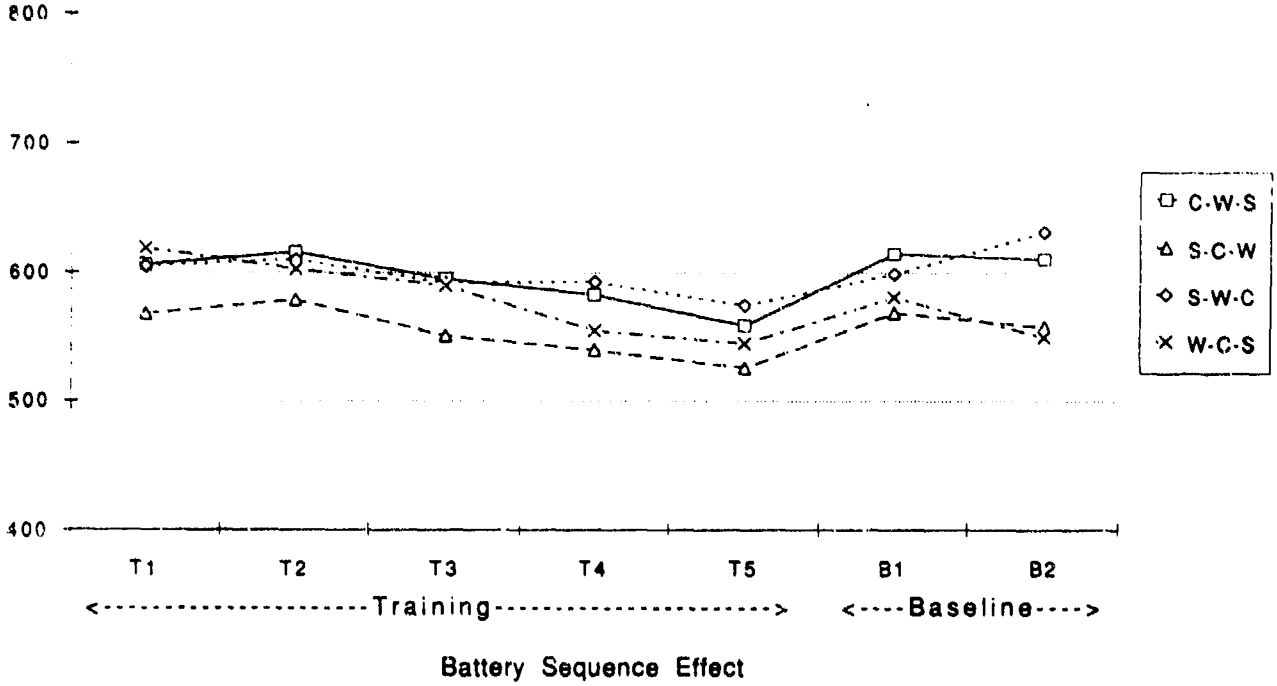
STRES Reaction Time - INVERT (5)
Percent Correct



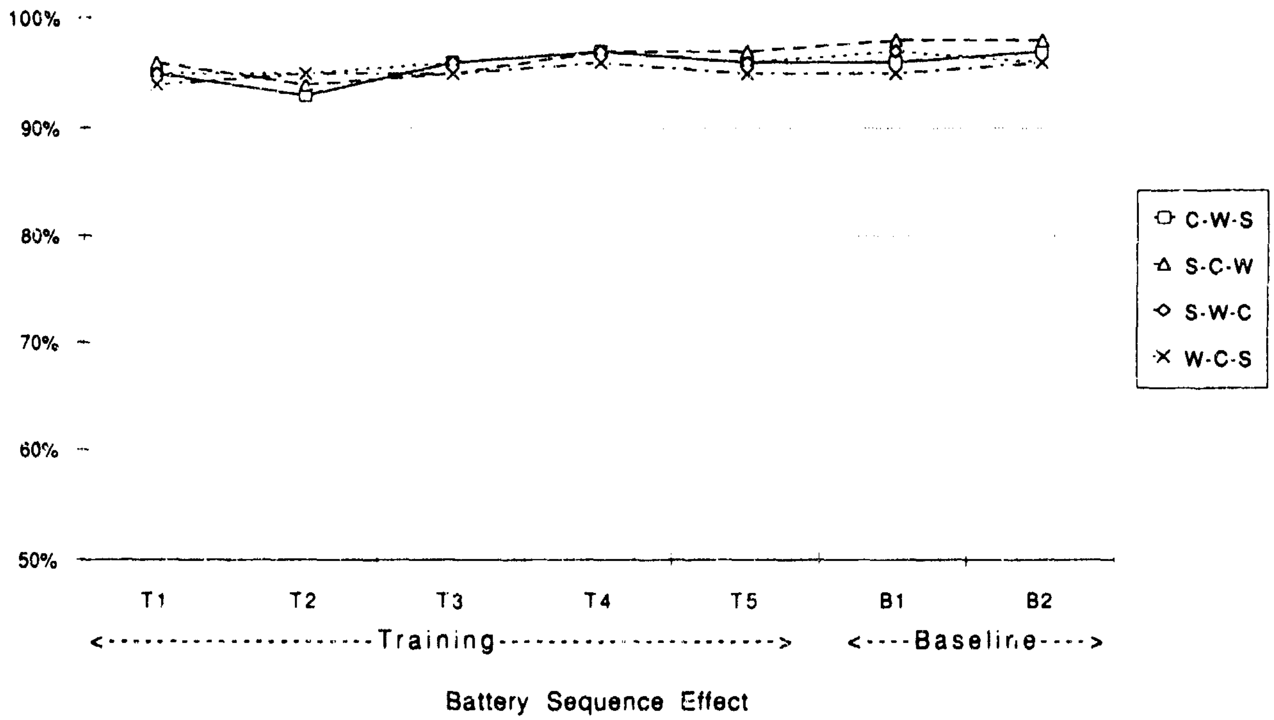
<-------------------Training-------------------> <----Baseline---->

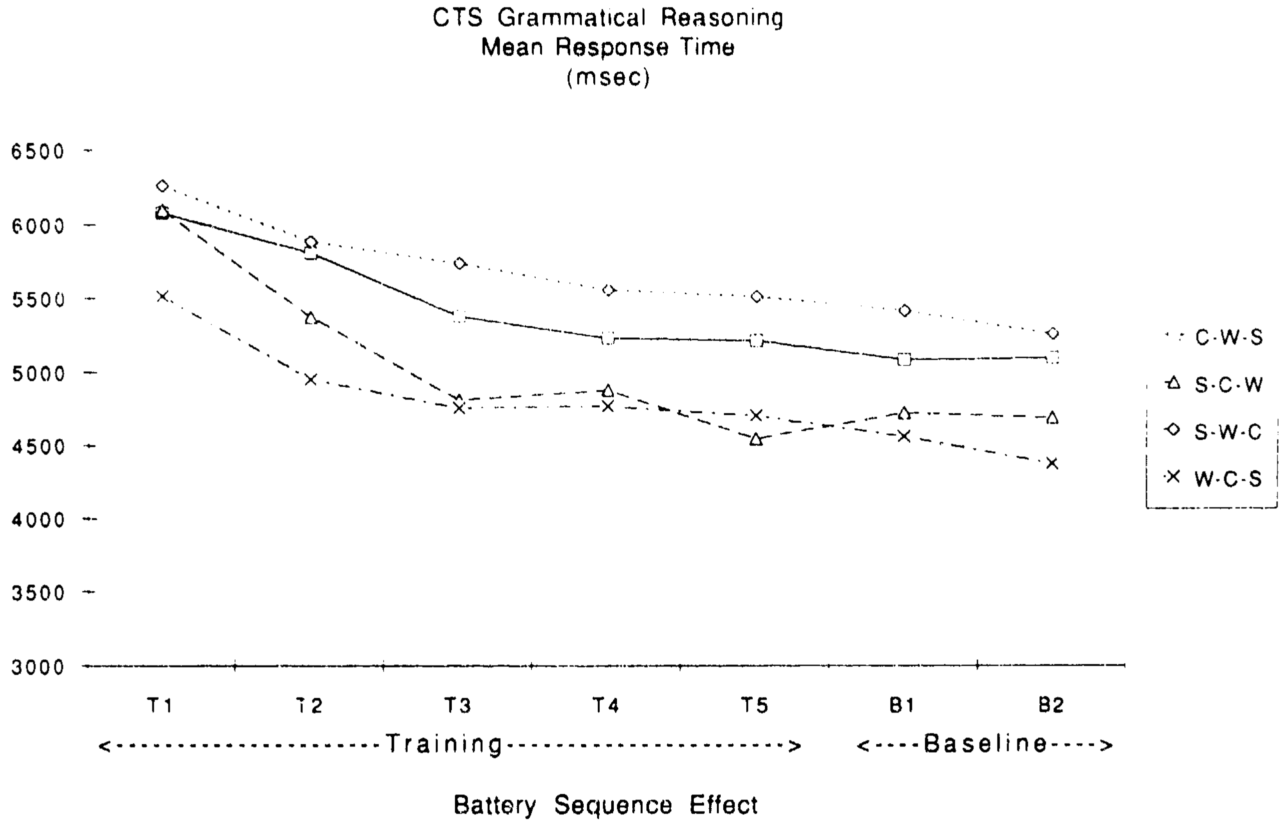
Battery Sequence Effect

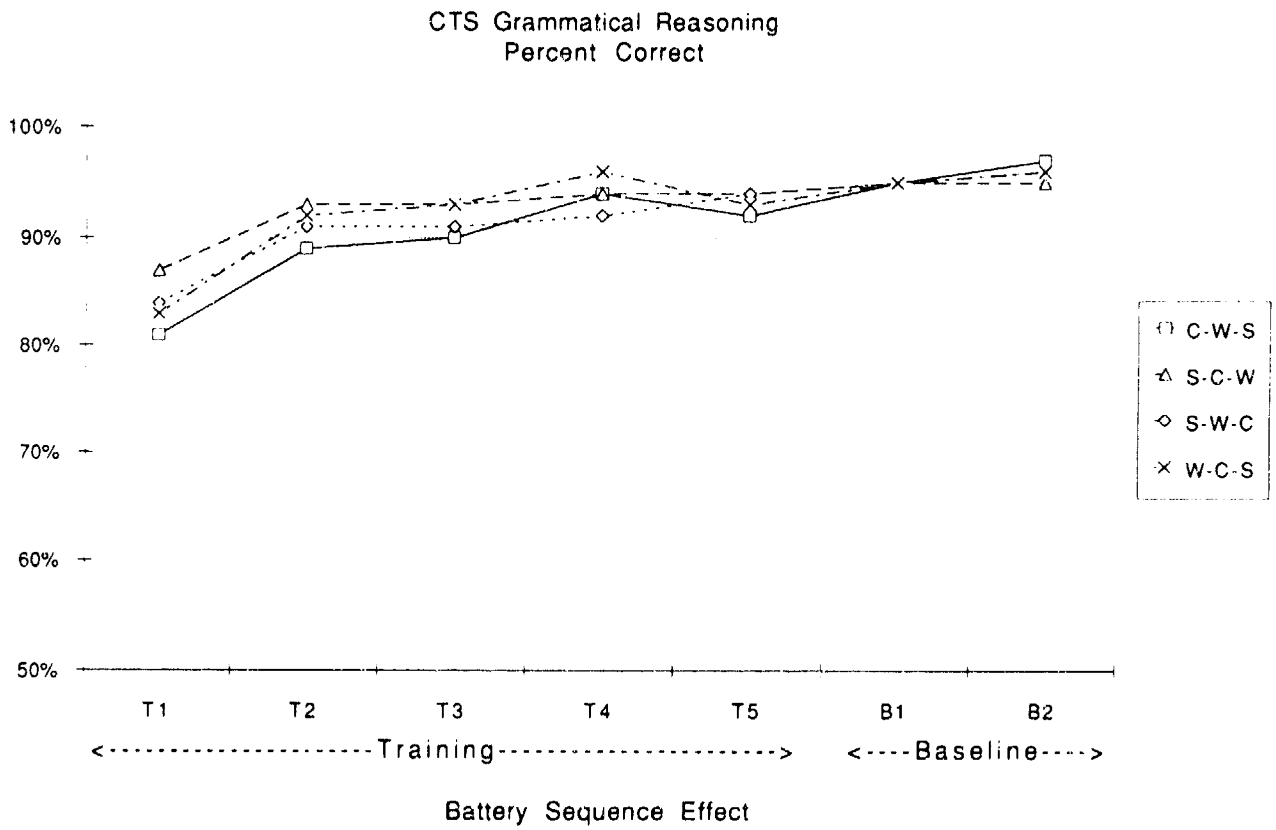
STRES Reaction Time - BASIC (6)
Mean Response Time
(msec)

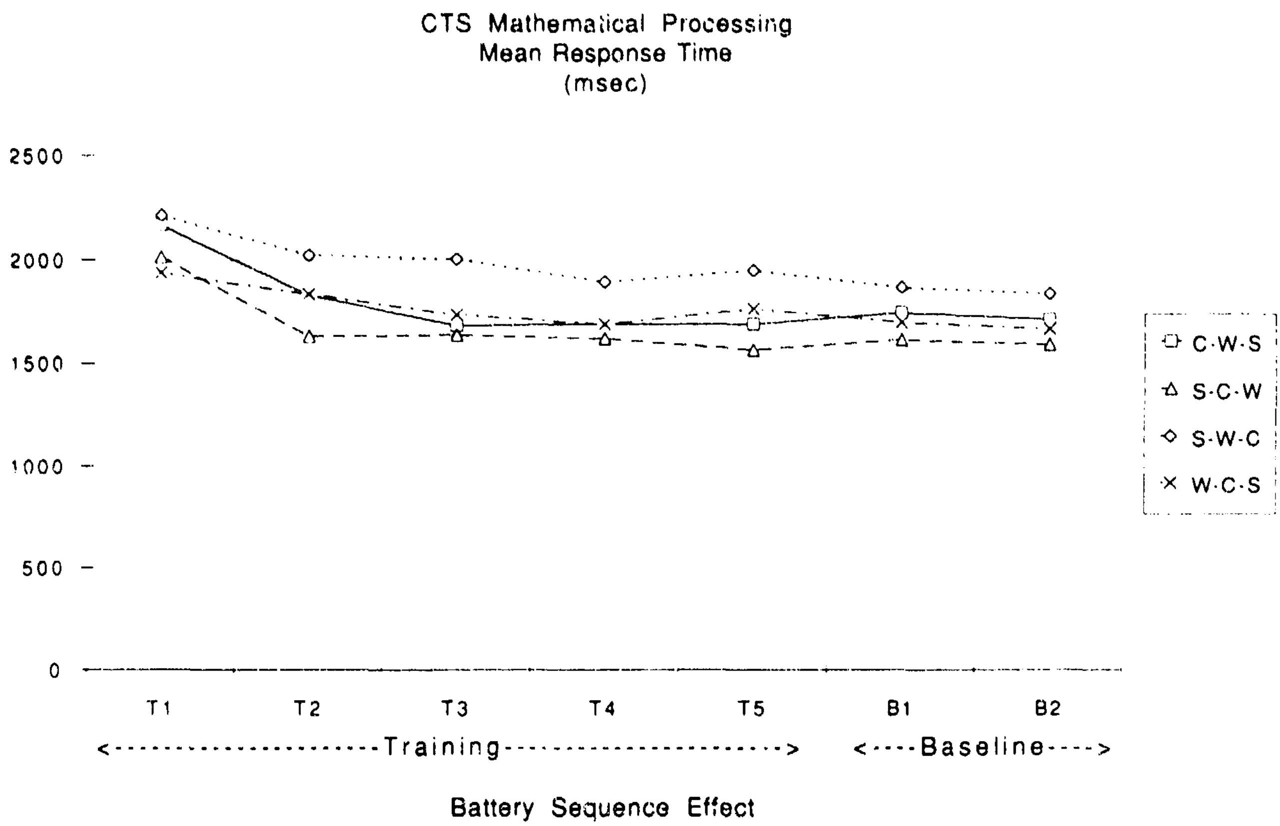


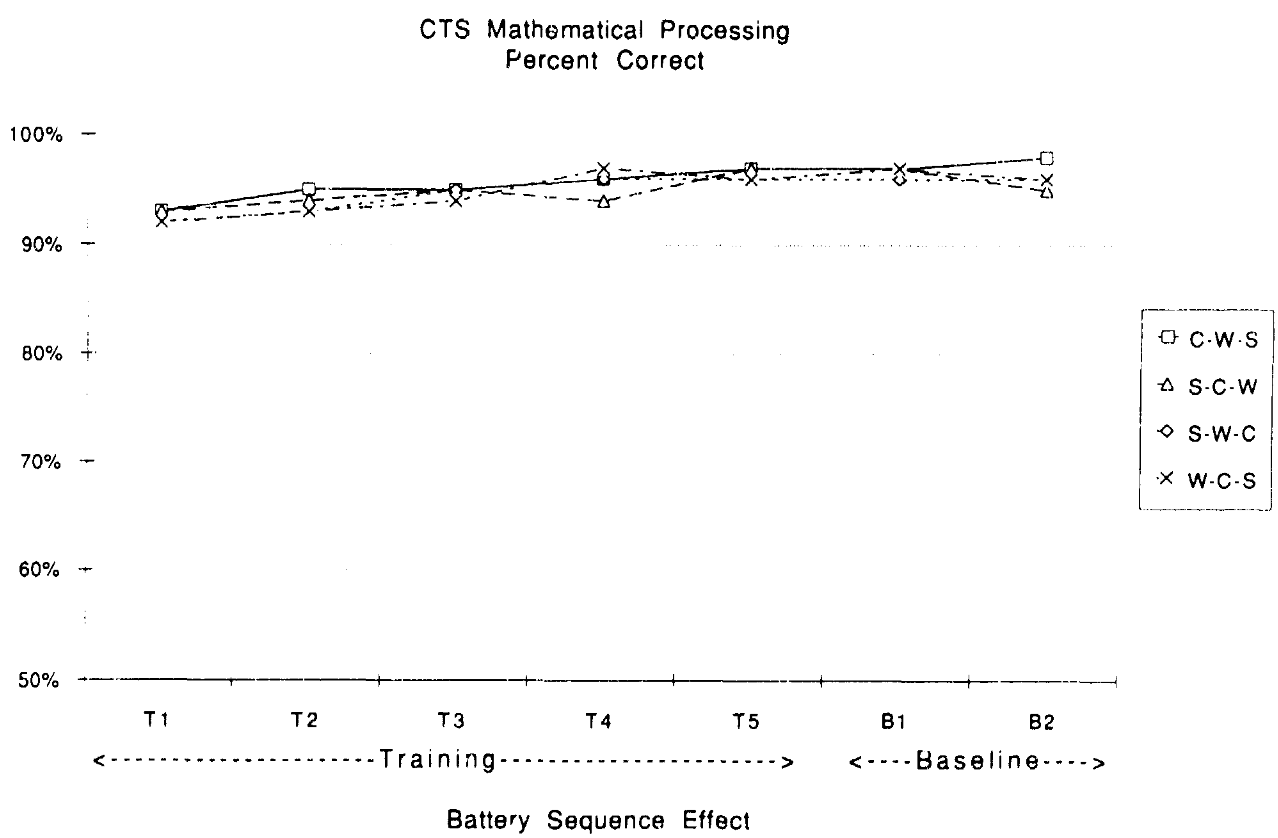
STRES Reaction Time - BASIC (6)
Percent Correct



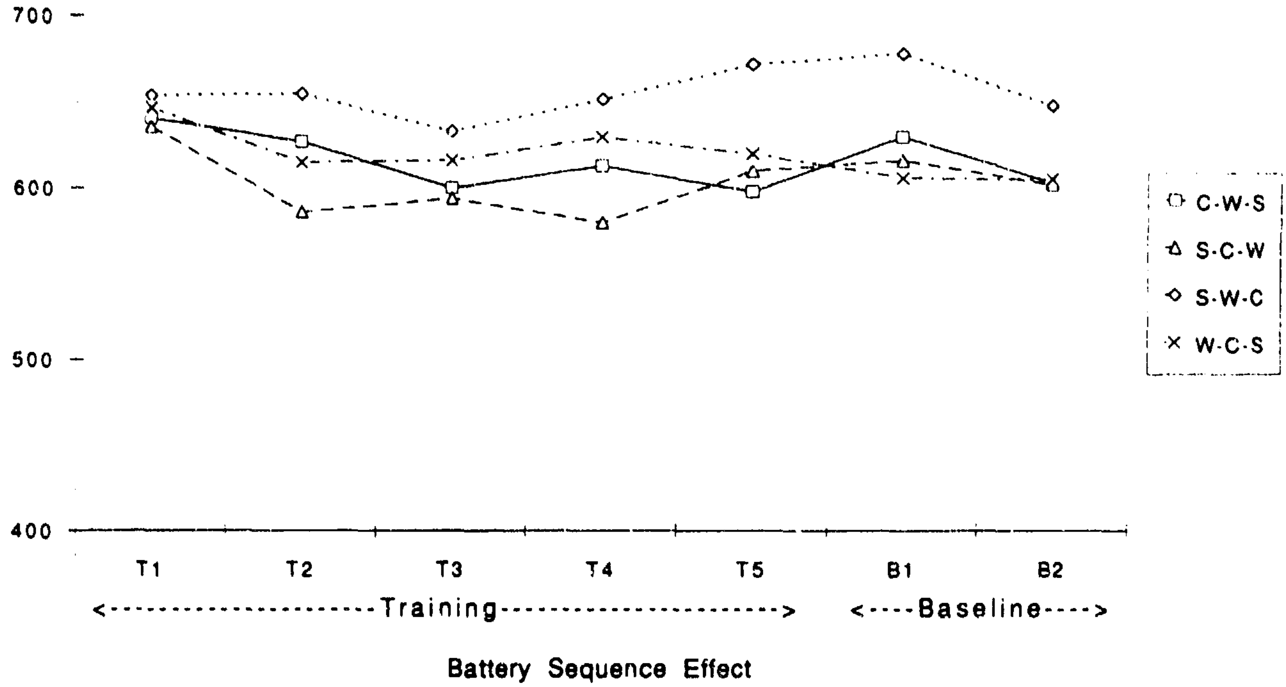

CTS Grammatical Reasoning
Mean Response Time
(msec)
6500
6000
5500
5000
4500
4000
3500
3000
T1
T2
T3
T4
T5
B1
B2
<----------------------Training----------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S


CTS Grammatical Reasoning
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<----------------------Training----------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

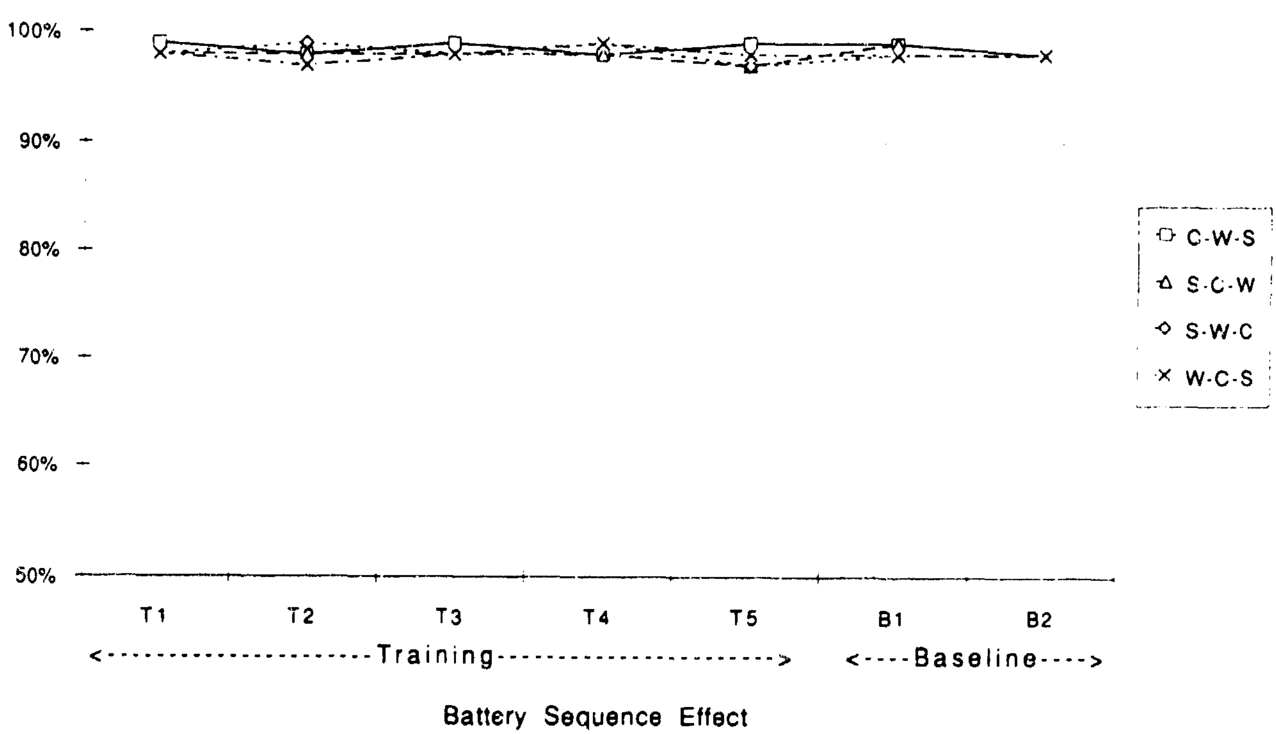

CTS Mathematical Processing
Mean Response Time
(msec)
2500
2000
1500
1000
500
0
T1
T2
T3
T4
T5
B1
B2
<--------------------Training--------------------> <----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

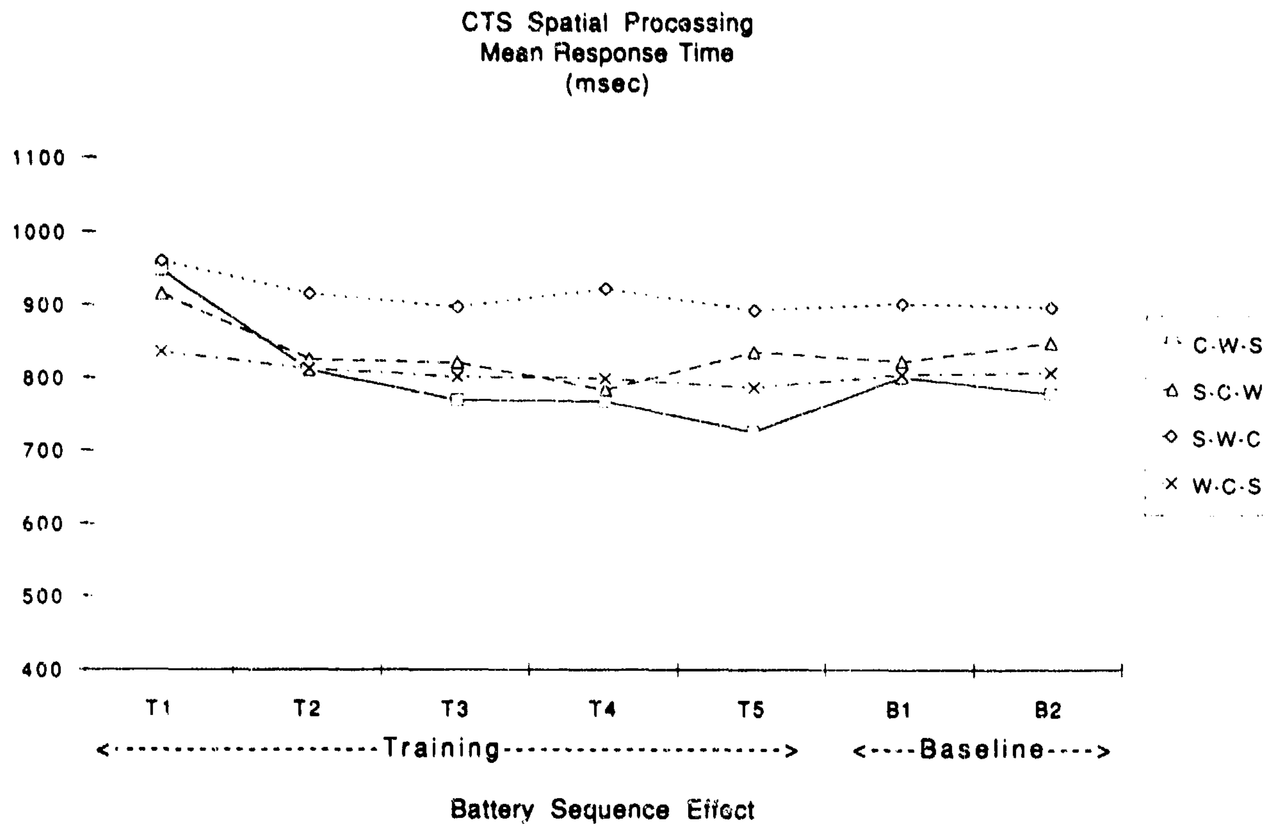

CTS Mathematical Processing
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<--------------------Training--------------------> <----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

CTS Memory Search-4
Mean Response Time
(msec)

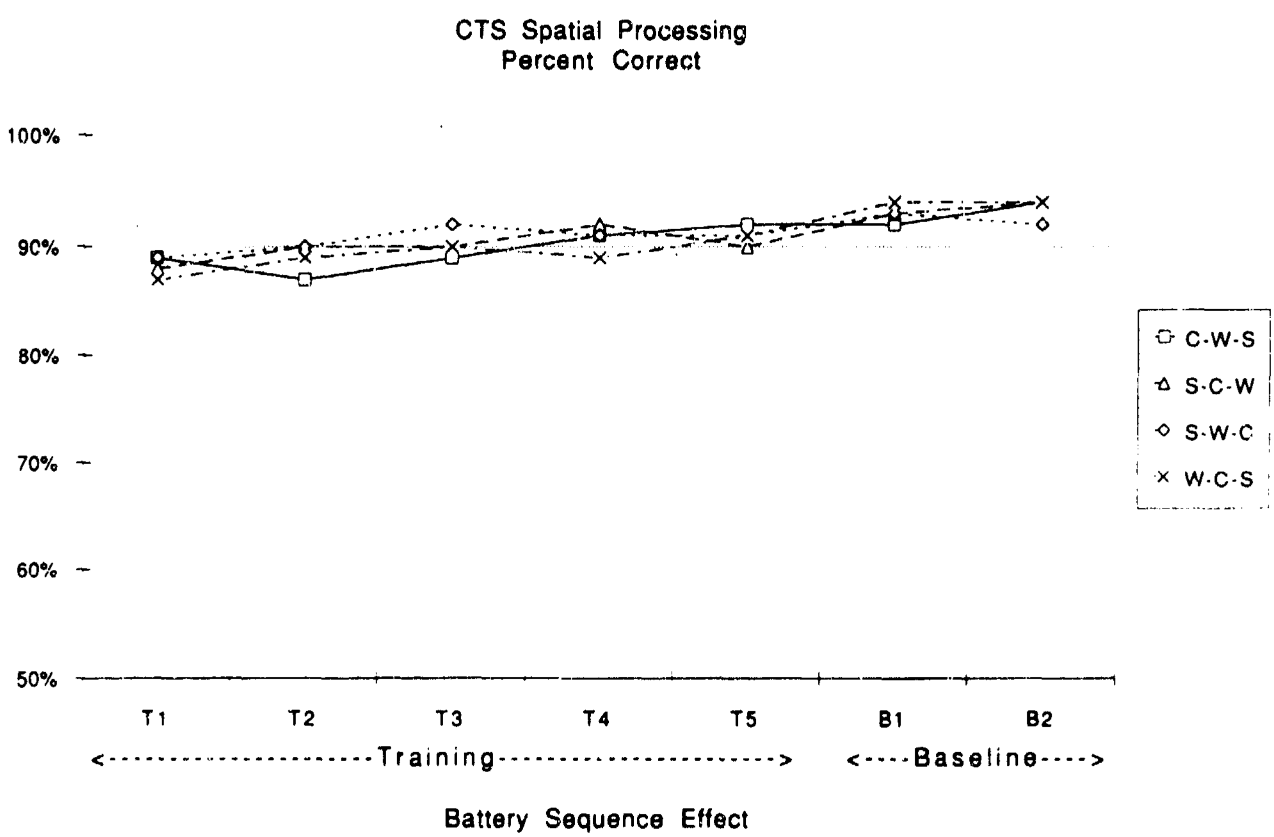


CTS Memory Search-4
Percent Correct



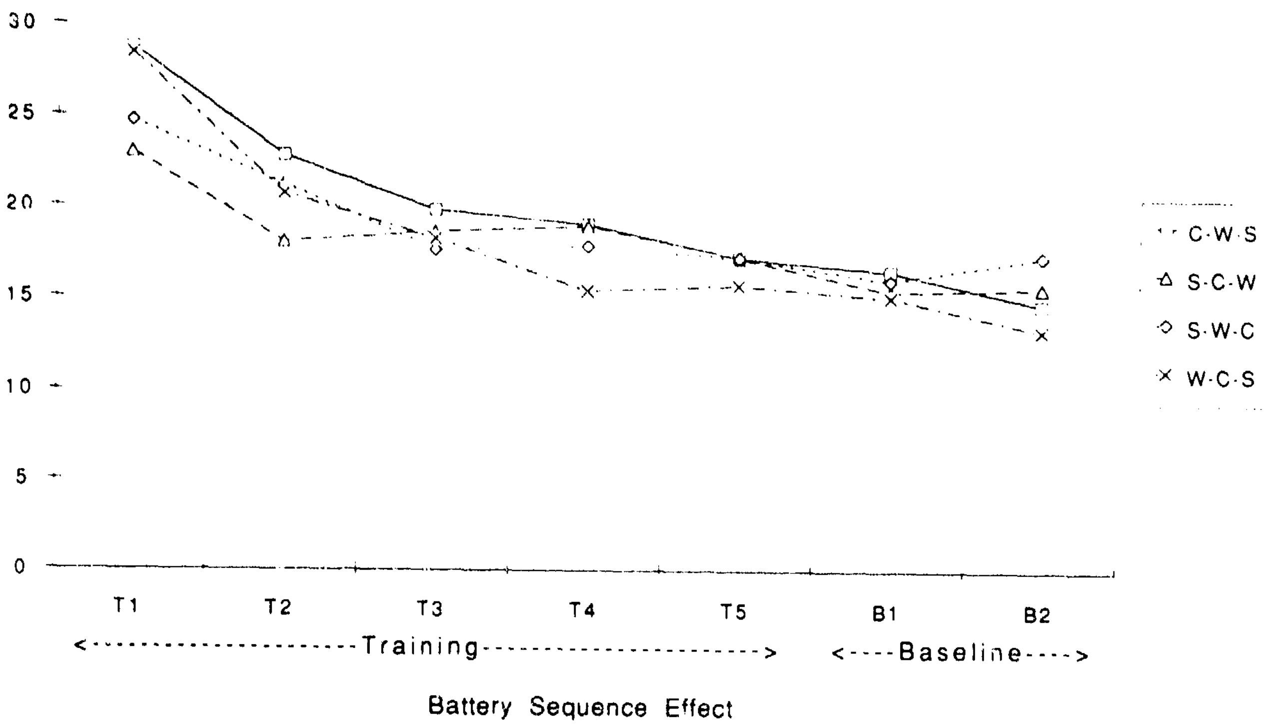


CTS Spatial Processing
Mean Response Time
(msec)
1100
1000
900
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<----------------------Training---------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

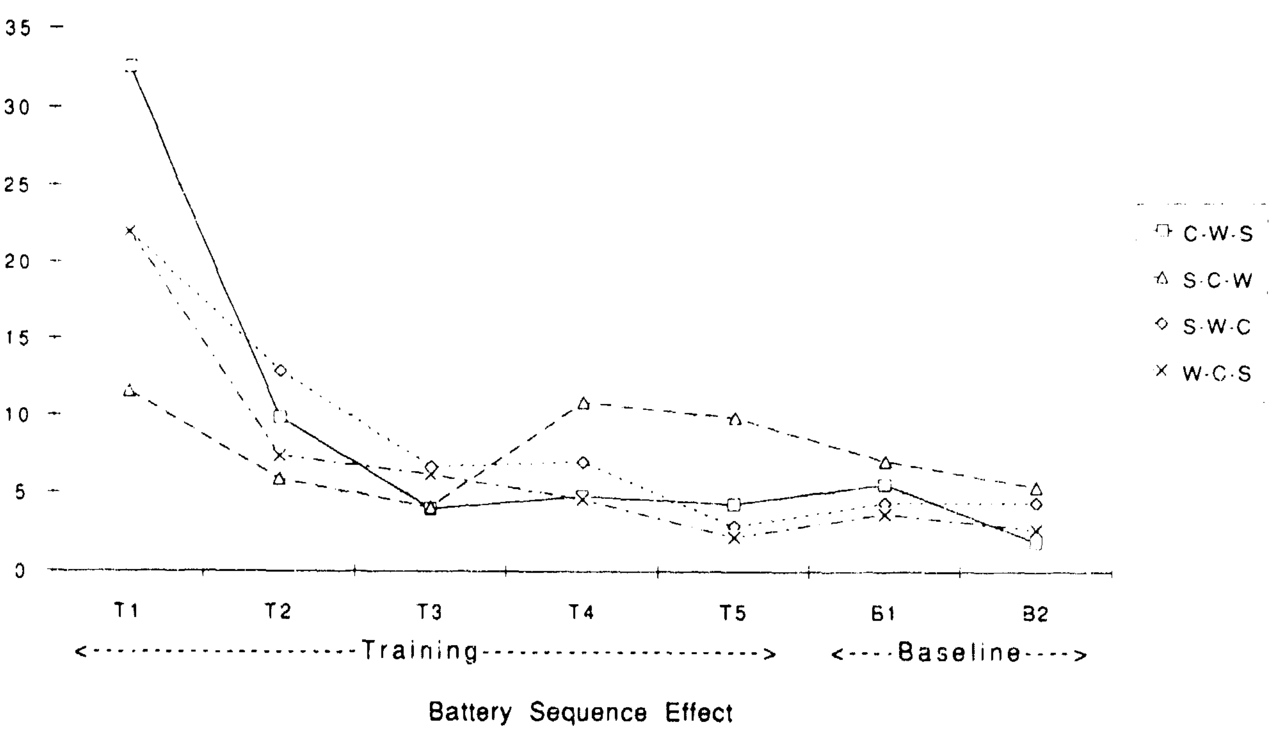


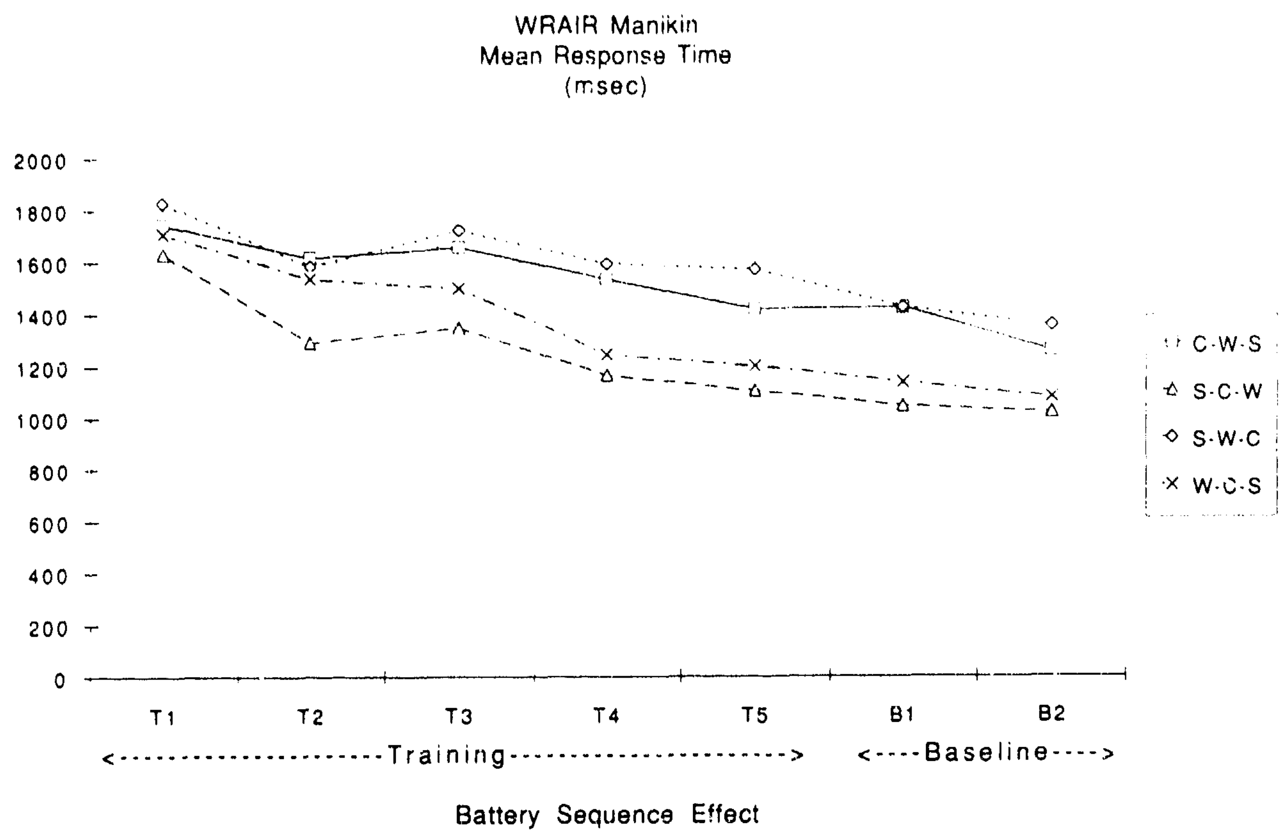
CTS Spatial Processing
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<----------------------Training---------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

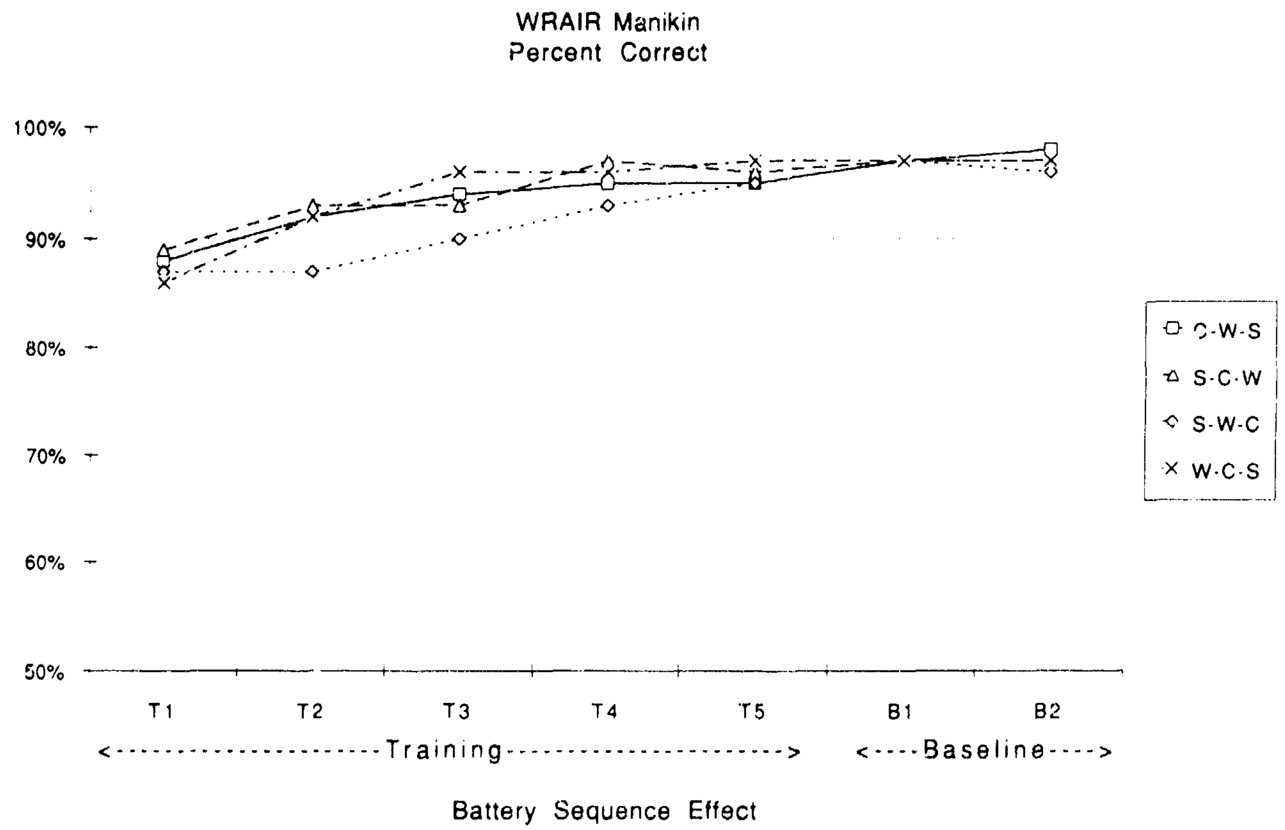
CTS Unstable Tracking
Edge Violations

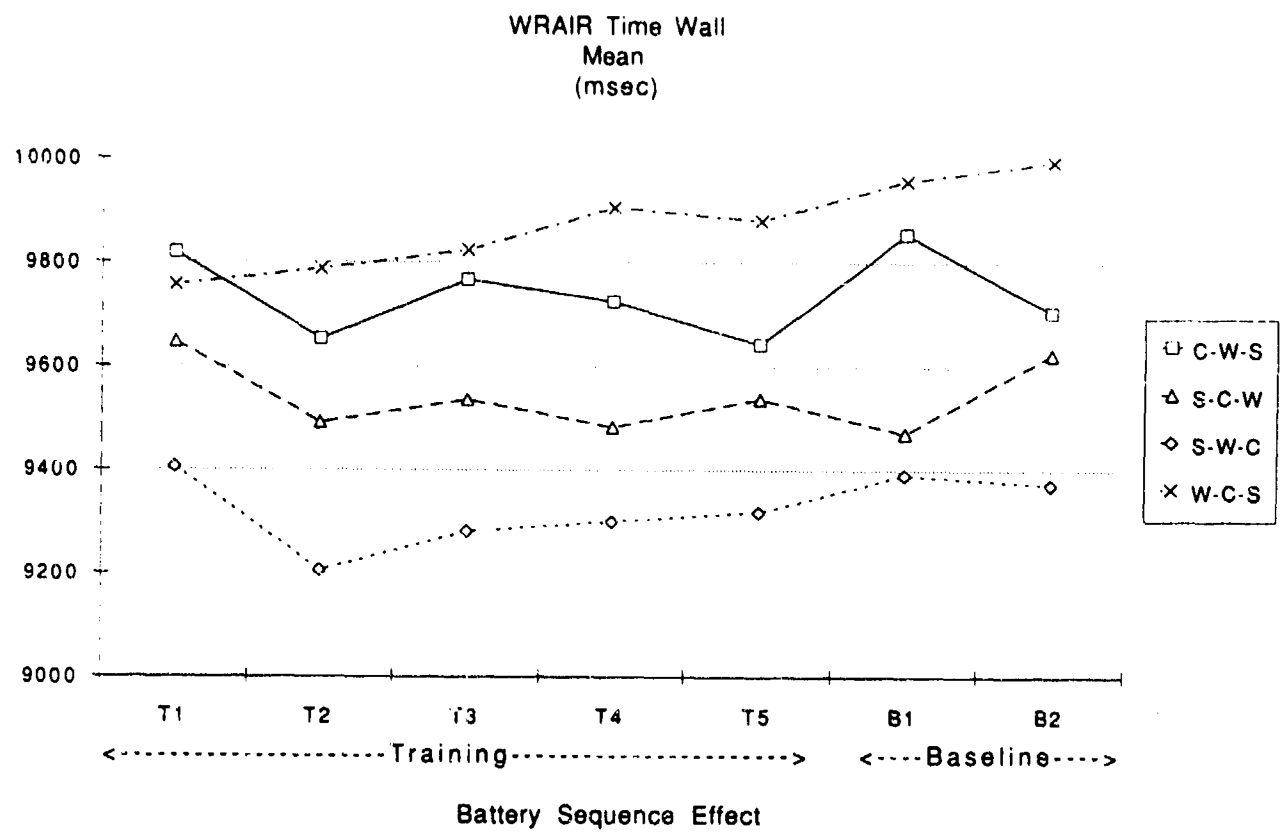


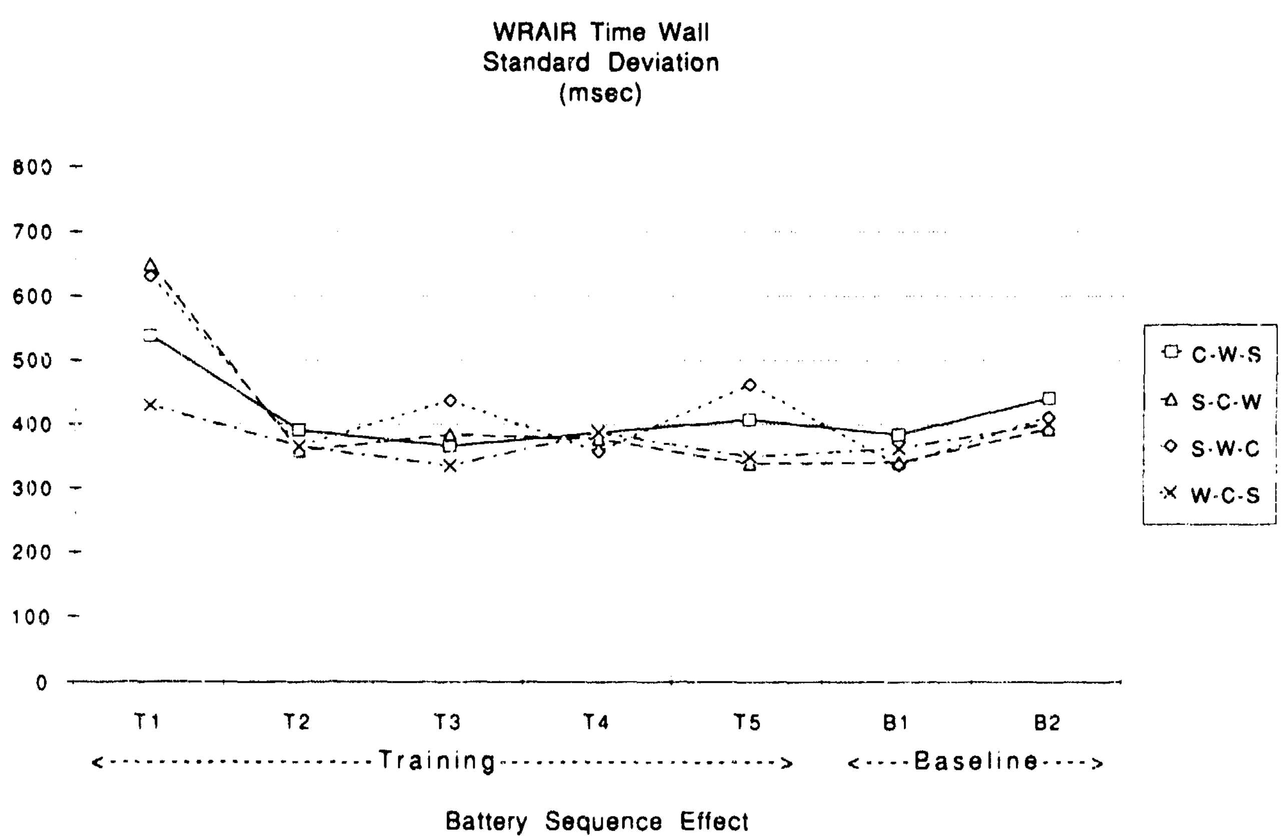
CTS Unstable Tracking
RMS Error



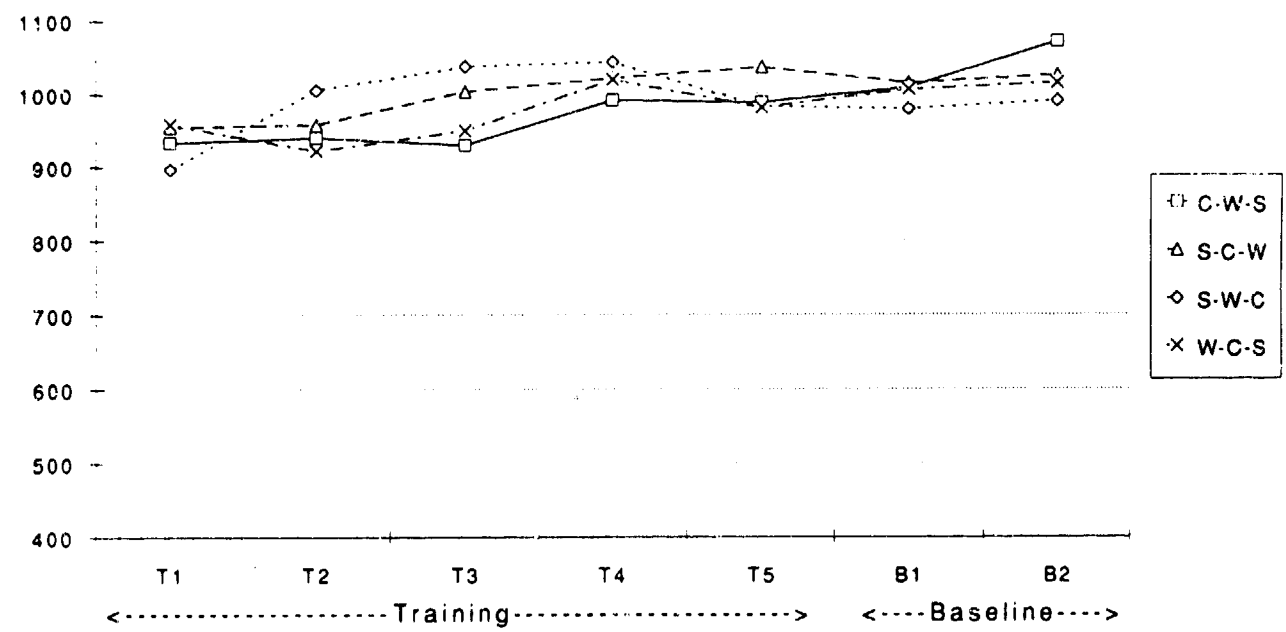

WRAIR Manikin
Mean Response Time
(msec)
2000
1800
1600
1400
1200
1000
800
600
400
200
0
T1
T2
T3
T4
T5
B1
B2
<--------------------Training-------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S


WRAIR Manikin
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<--------------------Training-------------------->
<----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S


WRAIR Time Wall
Mean
(msec)
10000
9800
9600
9400
9200
9000
T1
T2
T3
T4
T5
B1
B2
<--------------------Training-----------------------> <----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

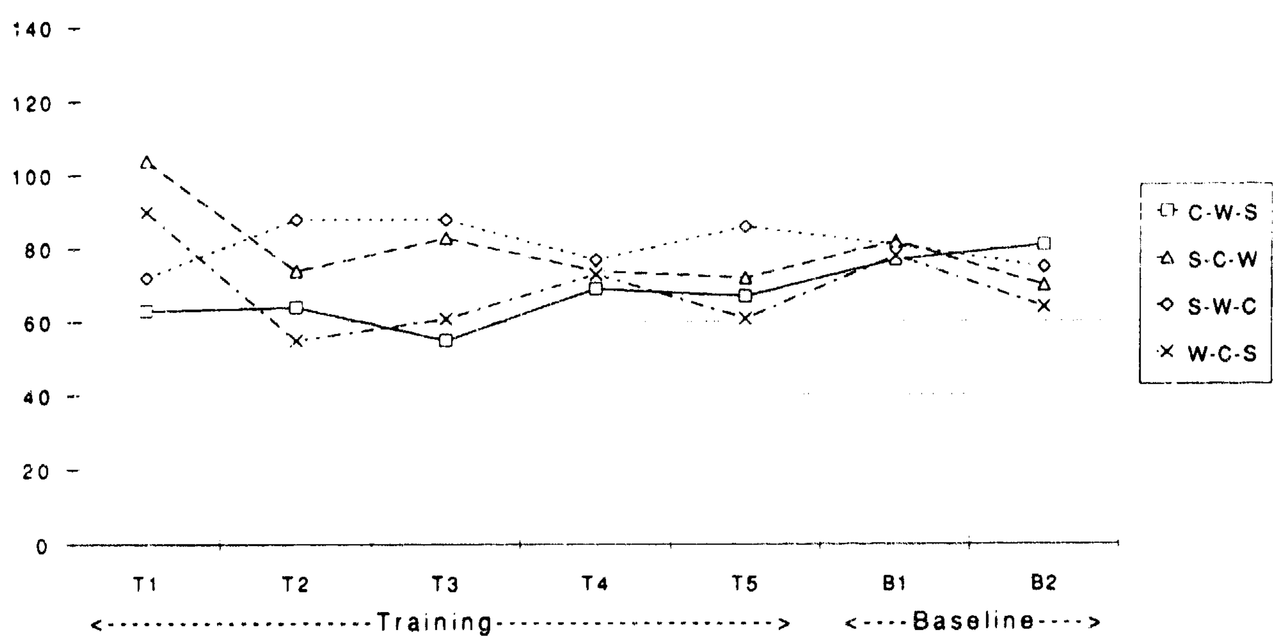

WRAIR Time Wall
Standard Deviation
(msec)
800
700
600
500
400
300
200
100
0
T1
T2
T3
T4
T5
B1
B2
<--------------------Training-----------------------> <----Baseline---->
Battery Sequence Effect
C-W-S
S-C-W
S-W-C
W-C-S

WRAIR Interval Production
Mean
(msec)



Battery Sequence Effect

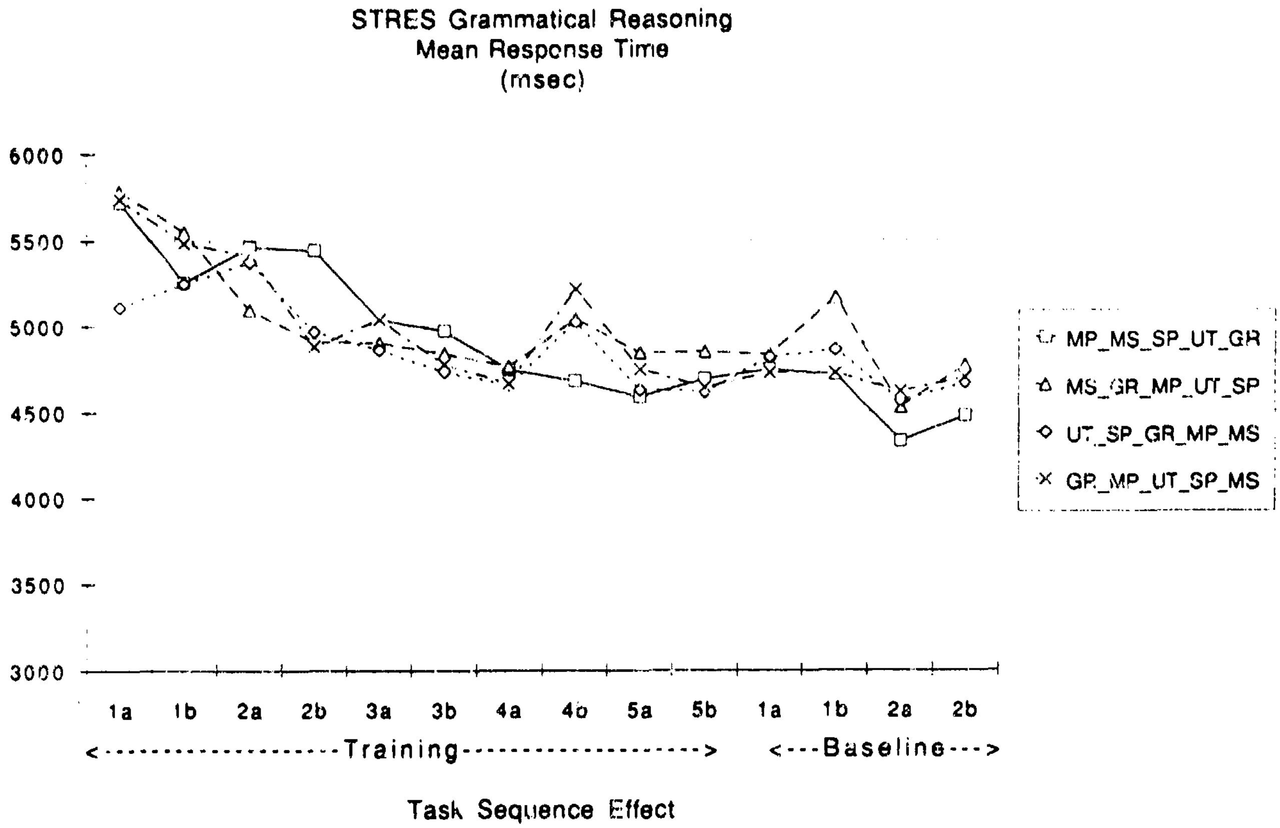
WRAIR Interval Production
Standard Deviation
(msec)

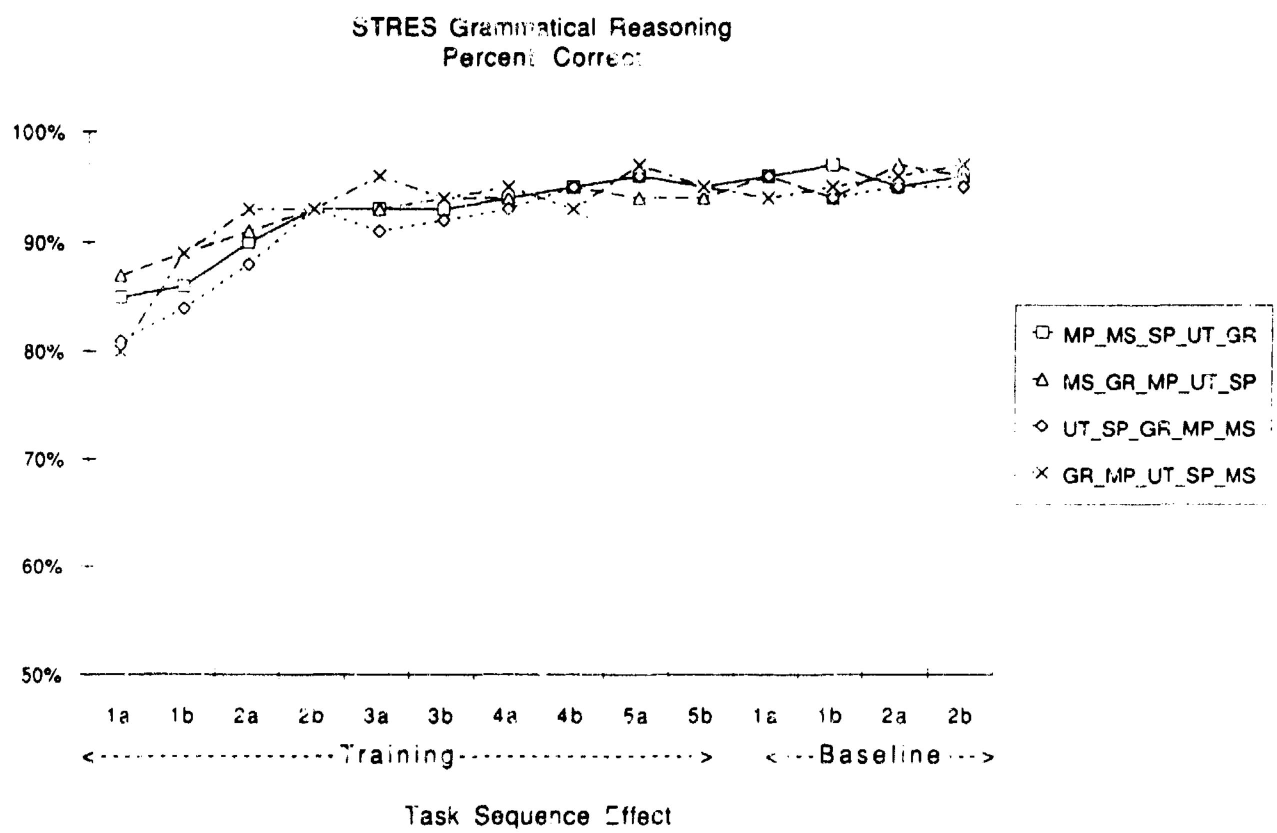


Battery Sequence Effect

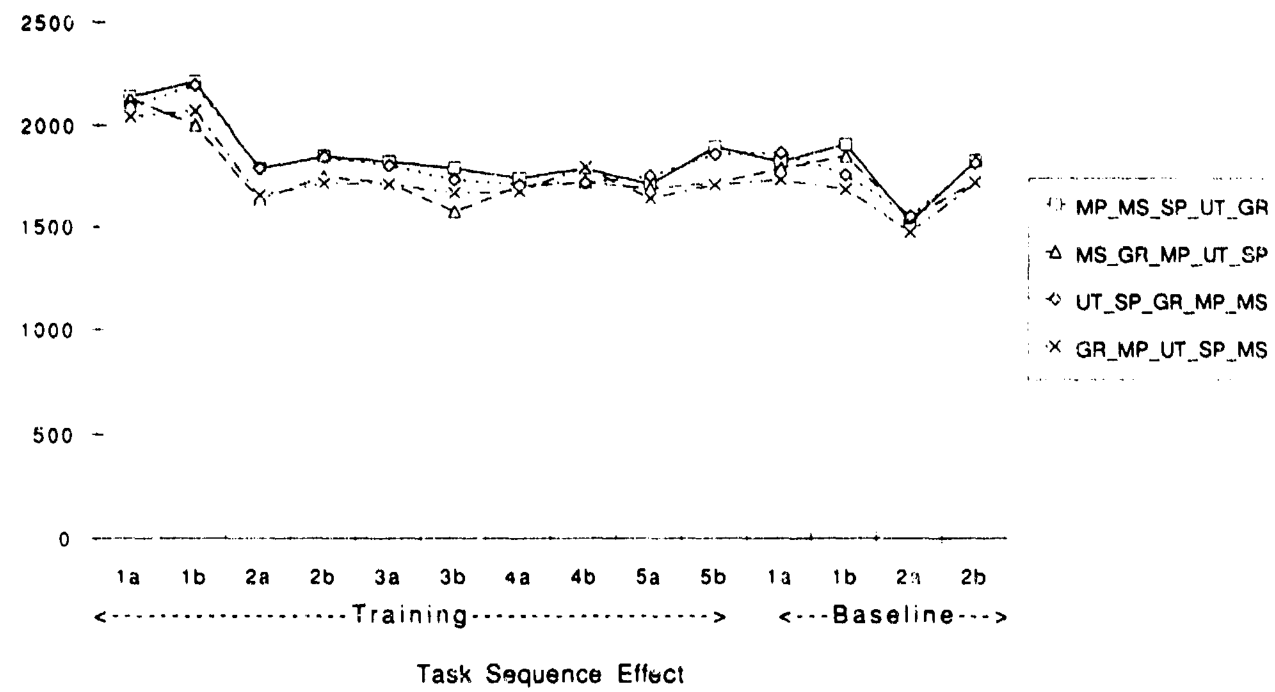
APPENDIX K

TASK ORDER DATA

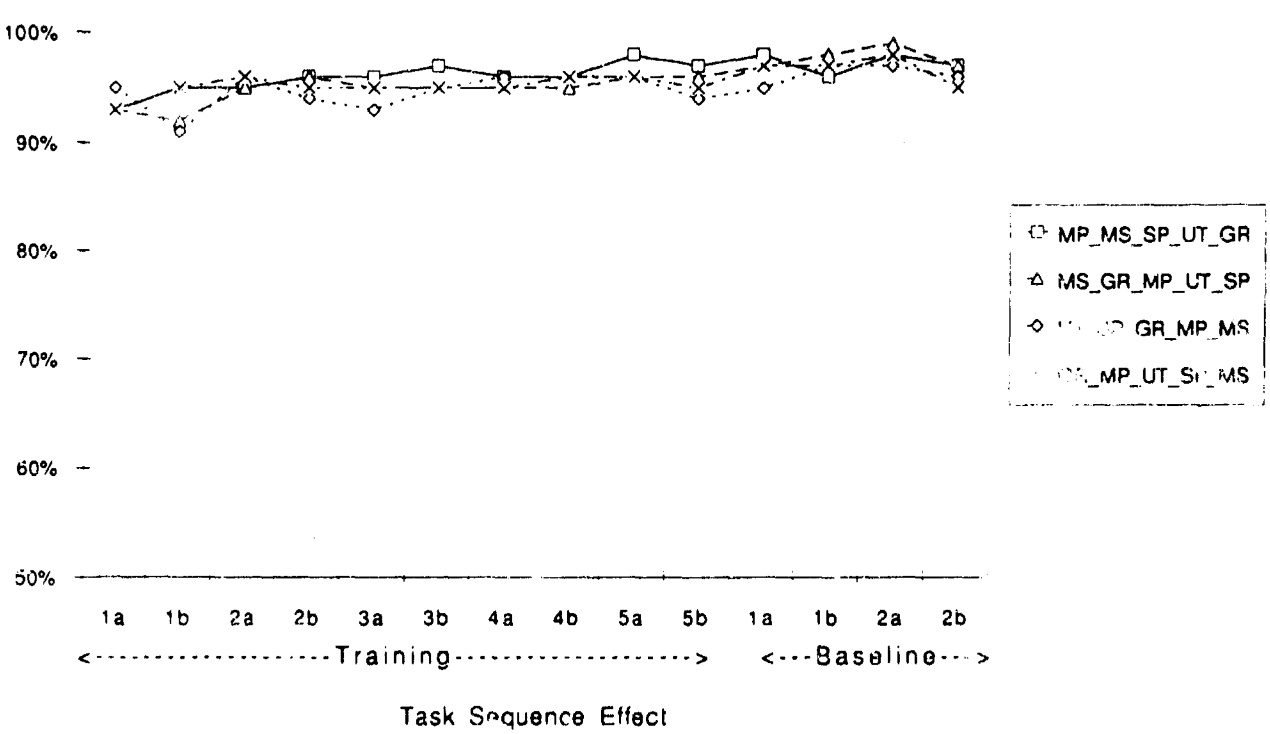

STRES Grammatical Reasoning
Mean Response Time
(msec)
6000
5500
5000
4500
4000
3500
3000
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<------------------Training------------------>
<---Baseline--->
Task Sequence Effect
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS

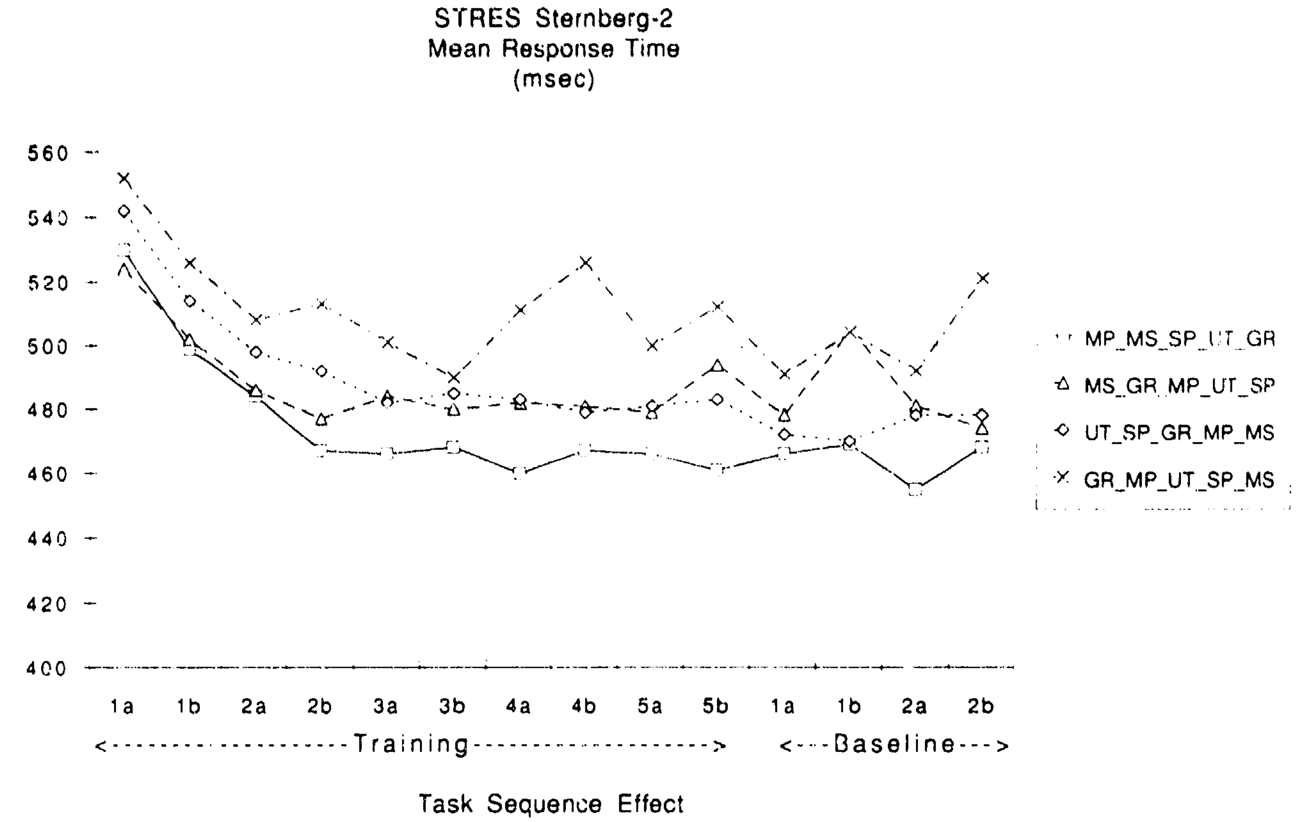

STRES Grammatical Reasoning
Percent Correct
100%
90%
80%
70%
60%
50%
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<------------------Training------------------>
<---Baseline--->
Task Sequence Effect
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS

STRES Mathematical Processing
Mean Response Time
(msec)

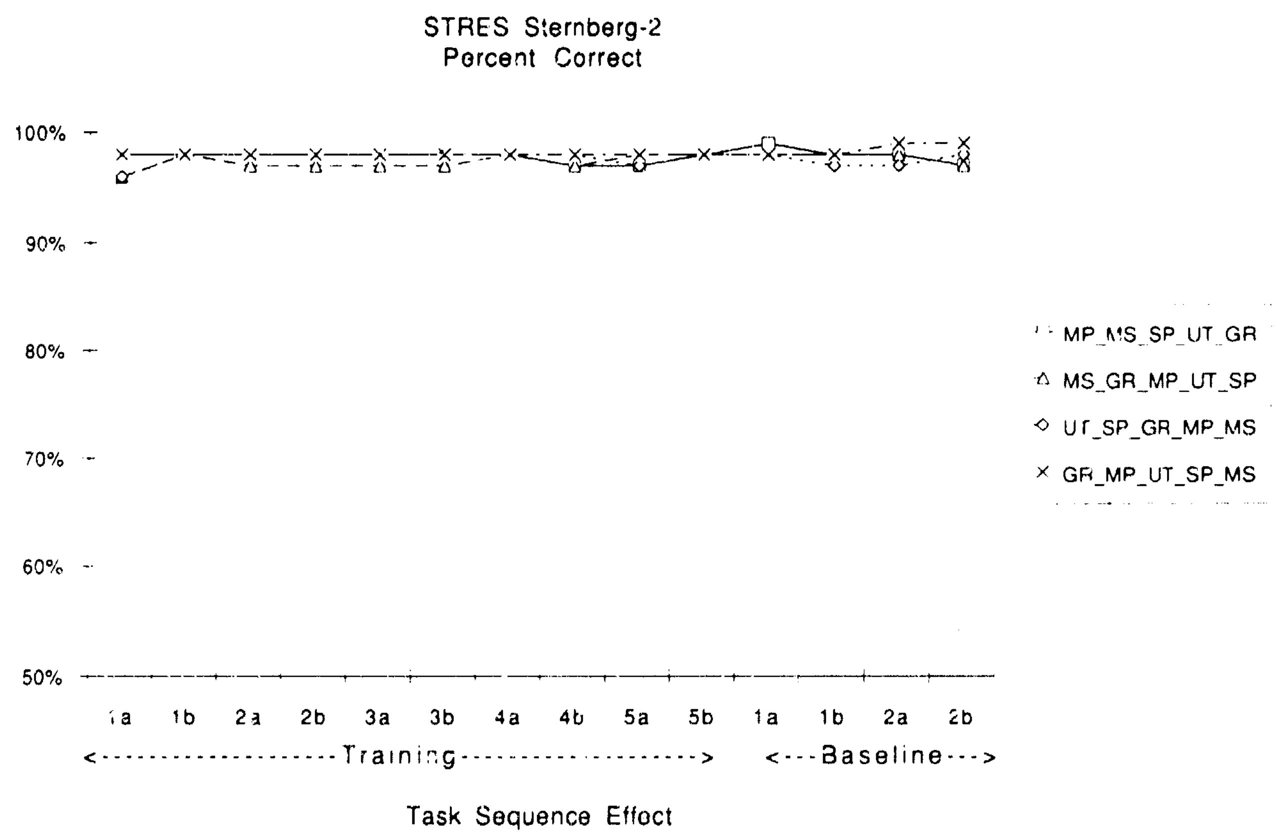


STRES Mathematical Processing
Percent Correct

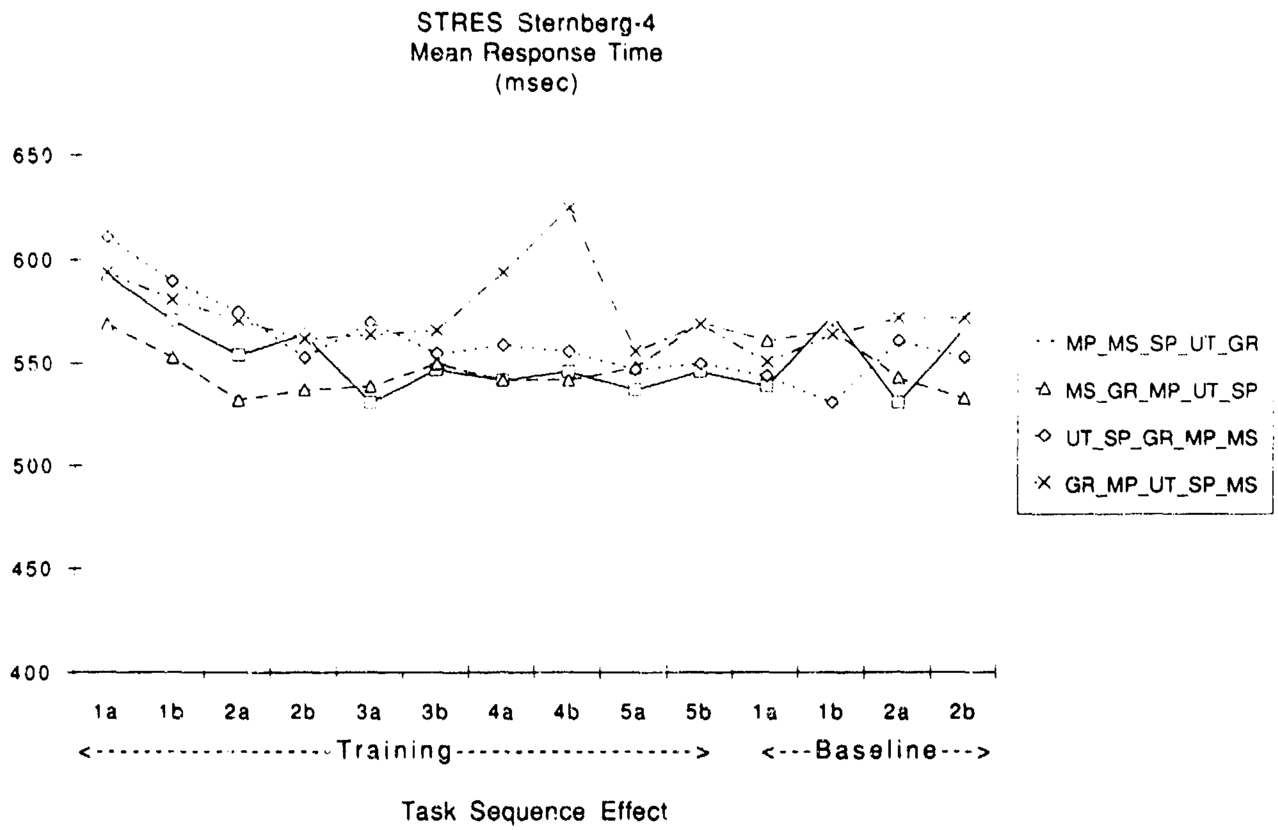


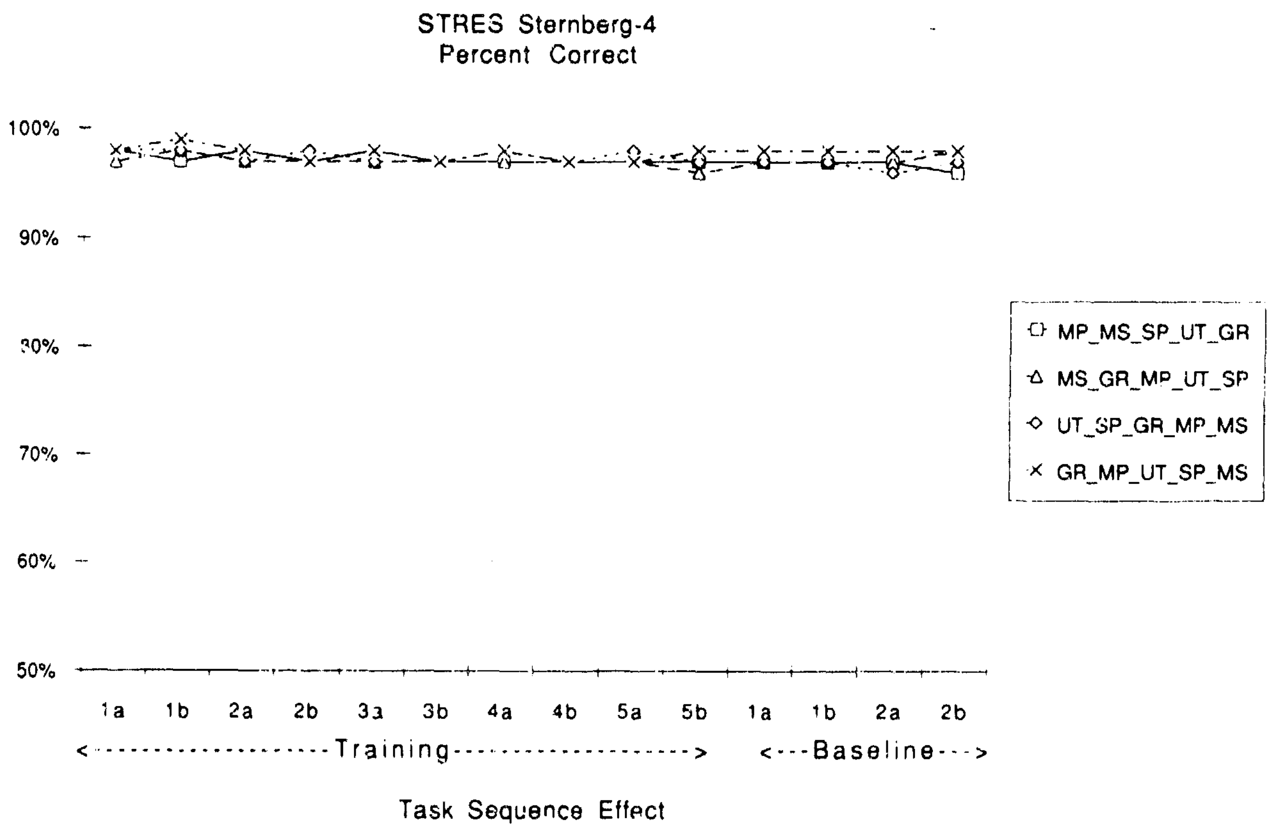


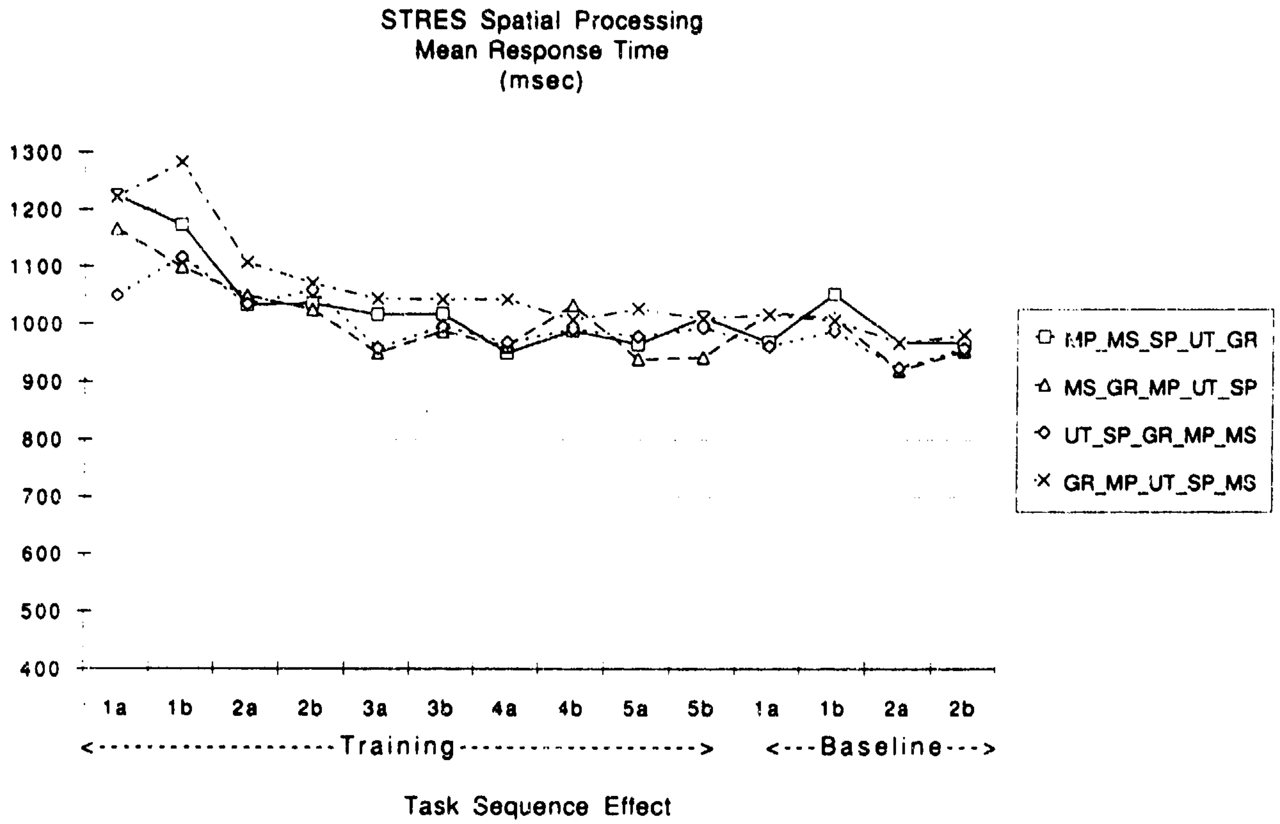
STRES Sternberg-2
Mean Response Time
(msec)
560
540
520
500
480
460
440
420
400
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<------------------------Training------------------------> <---Baseline--->
Task Sequence Effect
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS

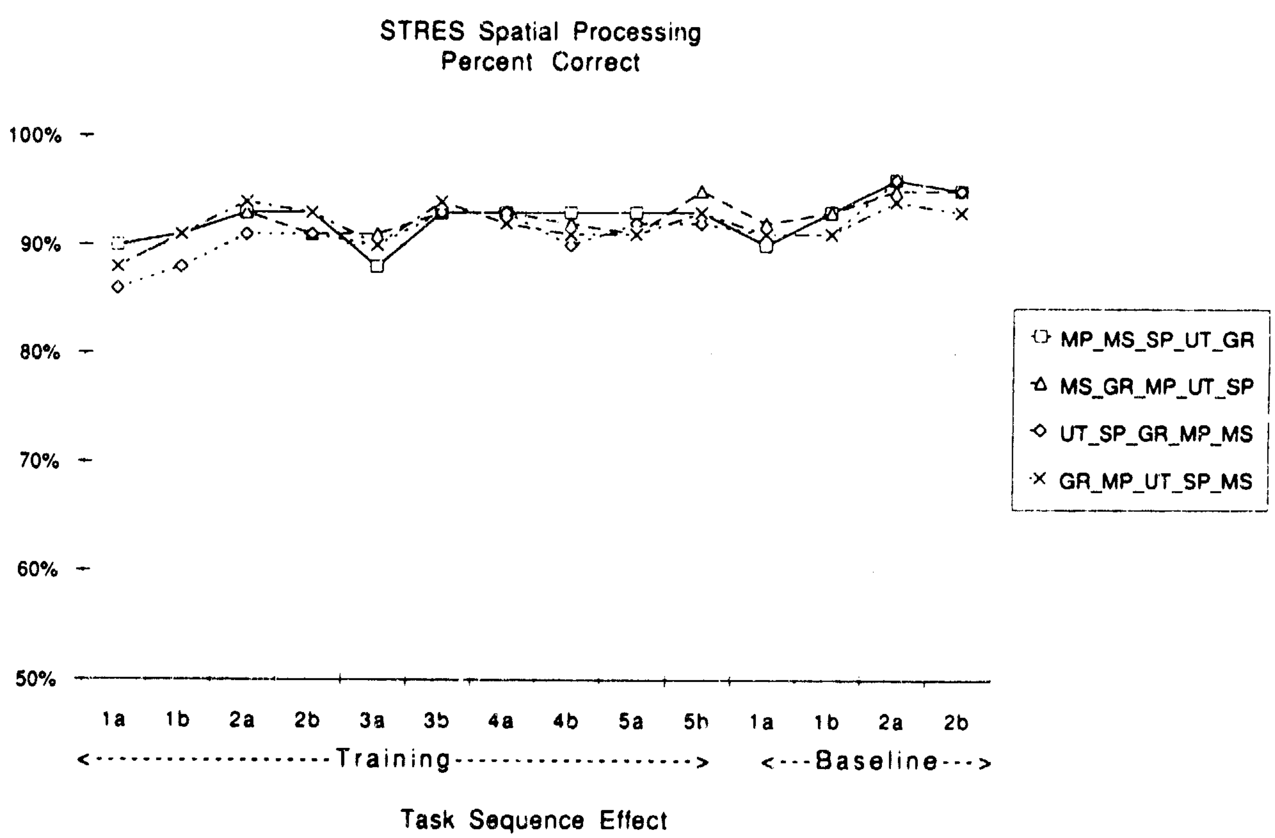


STRES Sternberg-2
Percent Correct
100%
90%
80%
70%
60%
50%
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<------------------------Training------------------------> <---Baseline--->
Task Sequence Effect
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS

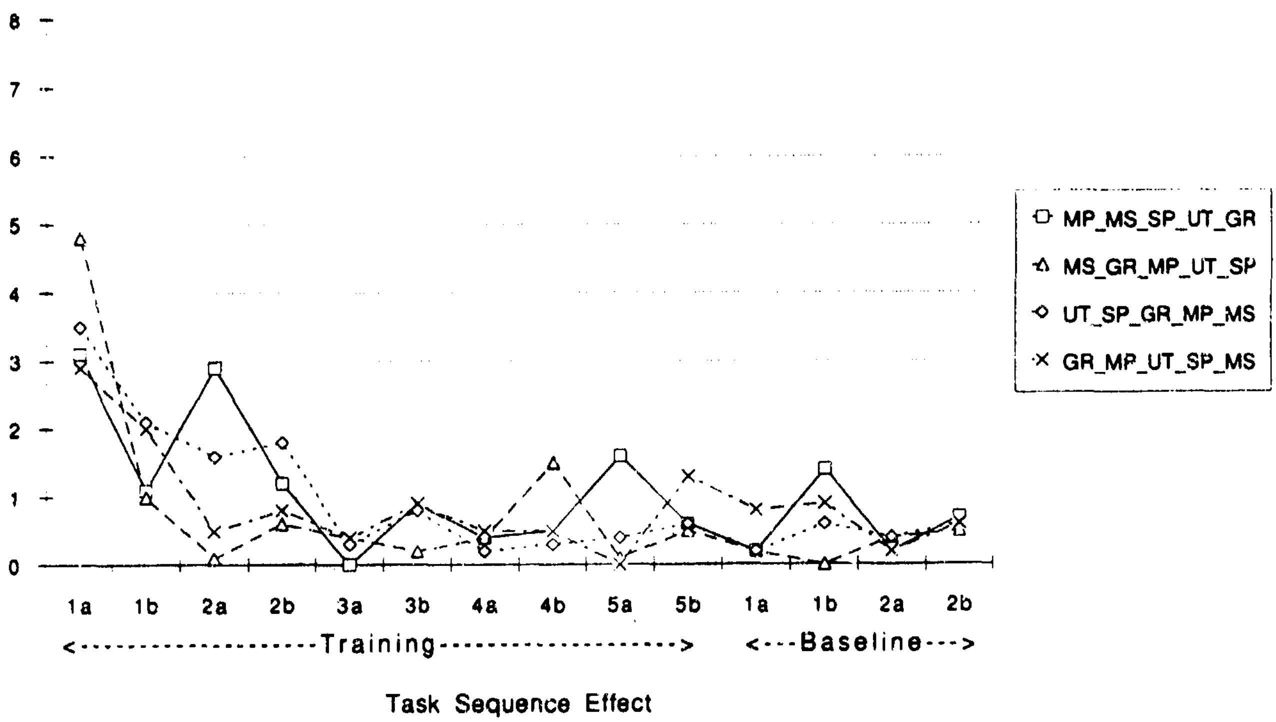

STRES Sternberg-4
Mean Response Time
(msec)
650
600
550
500
450
400
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<------------------Training------------------>
<---Baseline--->
Task Sequence Effect
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS


STRES Sternberg-4
Percent Correct
100%
90%
80%
70%
60%
50%
1a
1b
2a
2b
3a
3b
4a
4b
5a
5b
1a
1b
2a
2b
<------------------Training------------------>
<---Baseline--->
Task Sequence Effect
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS

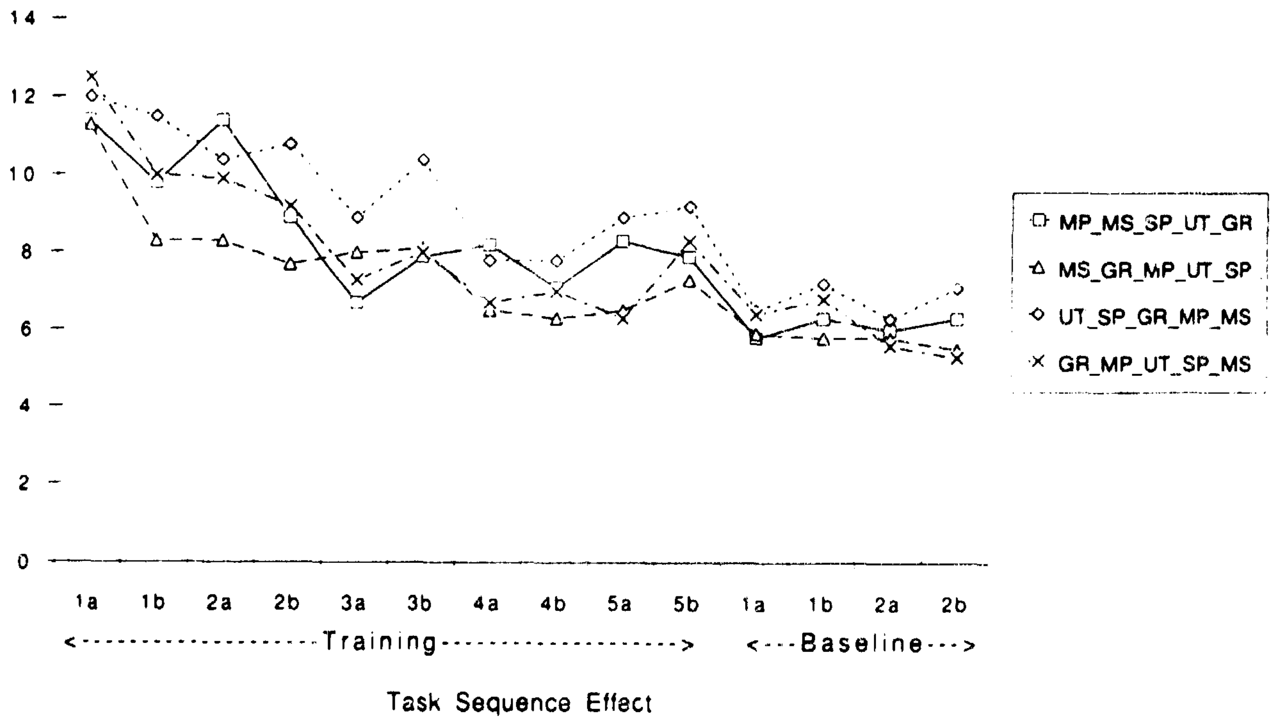

STRES Spatial Processing
Mean Response Time
(msec)
1300
1200
1100
1000
900
800
700
600
500
400
1a 1b 2a 2b 3a 3b 4a 4b 5a 5b 1a 1b 2a 2b
<-------------------Training------------------->
<---Baseline--->
Task Sequence Effect
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS


STRES Spatial Processing
Percent Correct
100%
90%
80%
70%
60%
50%
1a 1b 2a 2b 3a 3b 4a 4b 5a 5h 1a 1b 2a 2b
<-------------------Training------------------->
<---Baseline--->
Task Sequence Effect
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS

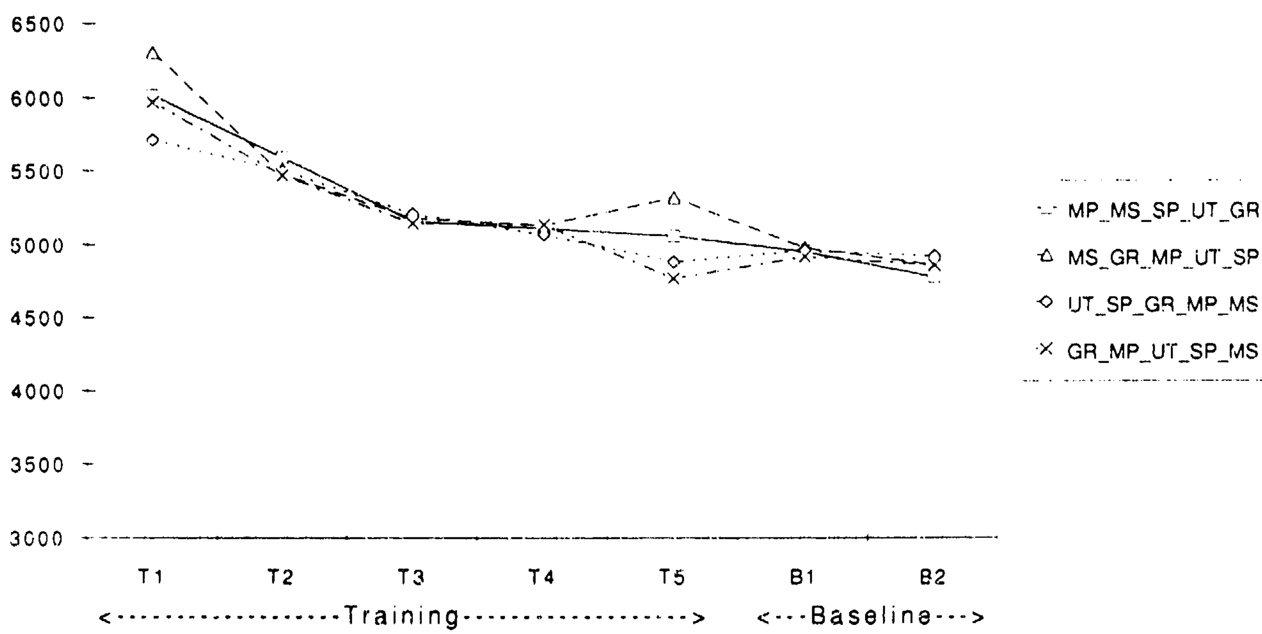
STRES Unstable Tracking
Edge Violations



STRES Unstable Tracking
RMS Error

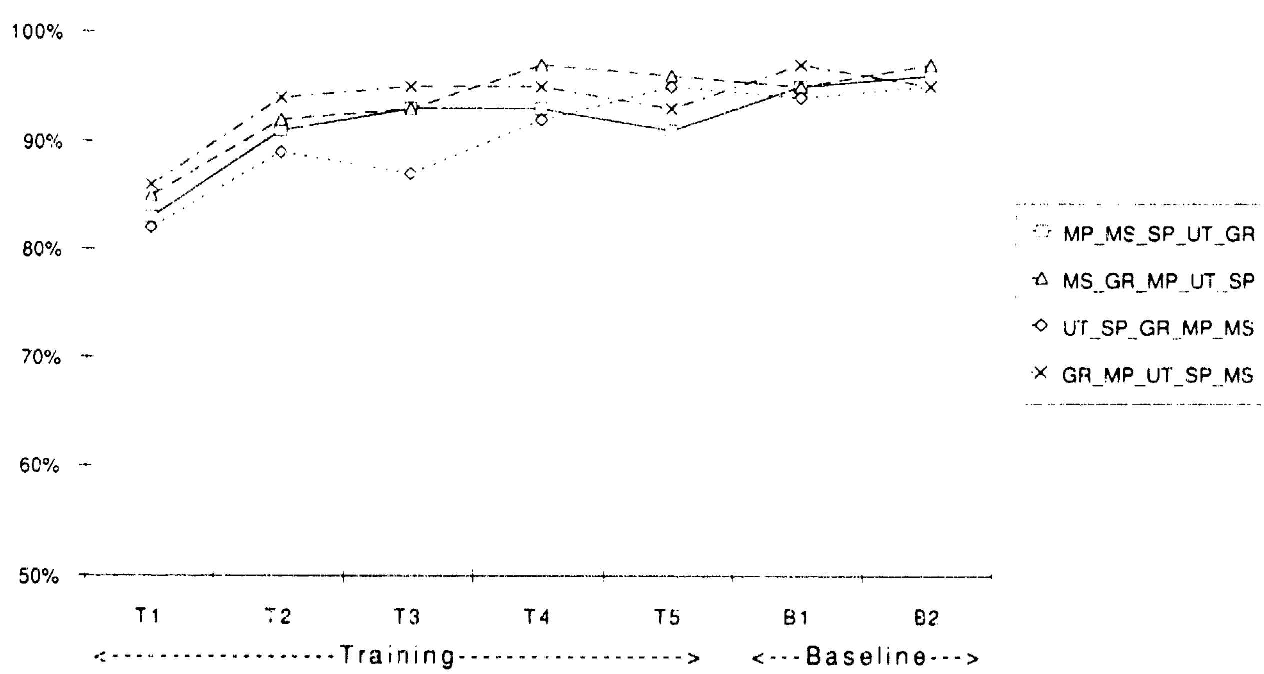


CTS Grammatical Reasoning
Mean Response Time
(msec)

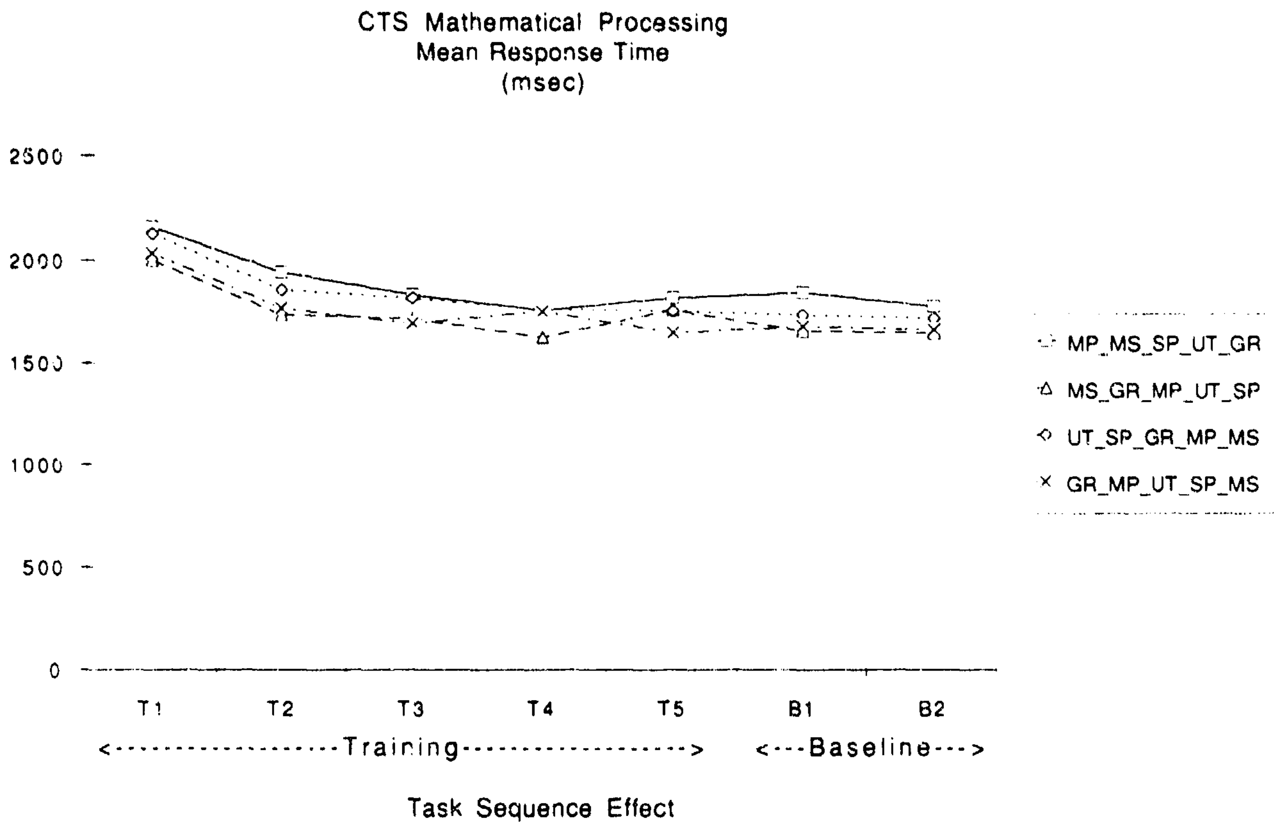


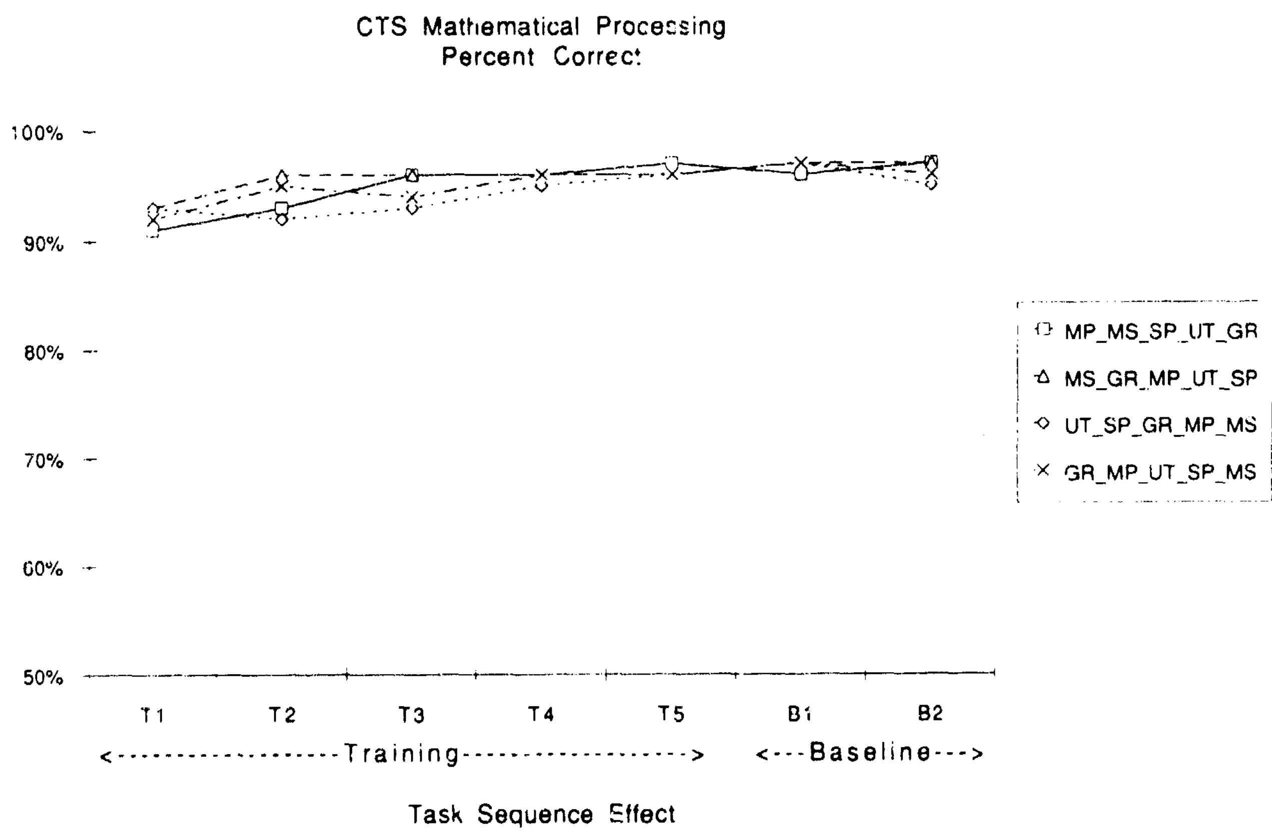
Task Sequence Effect

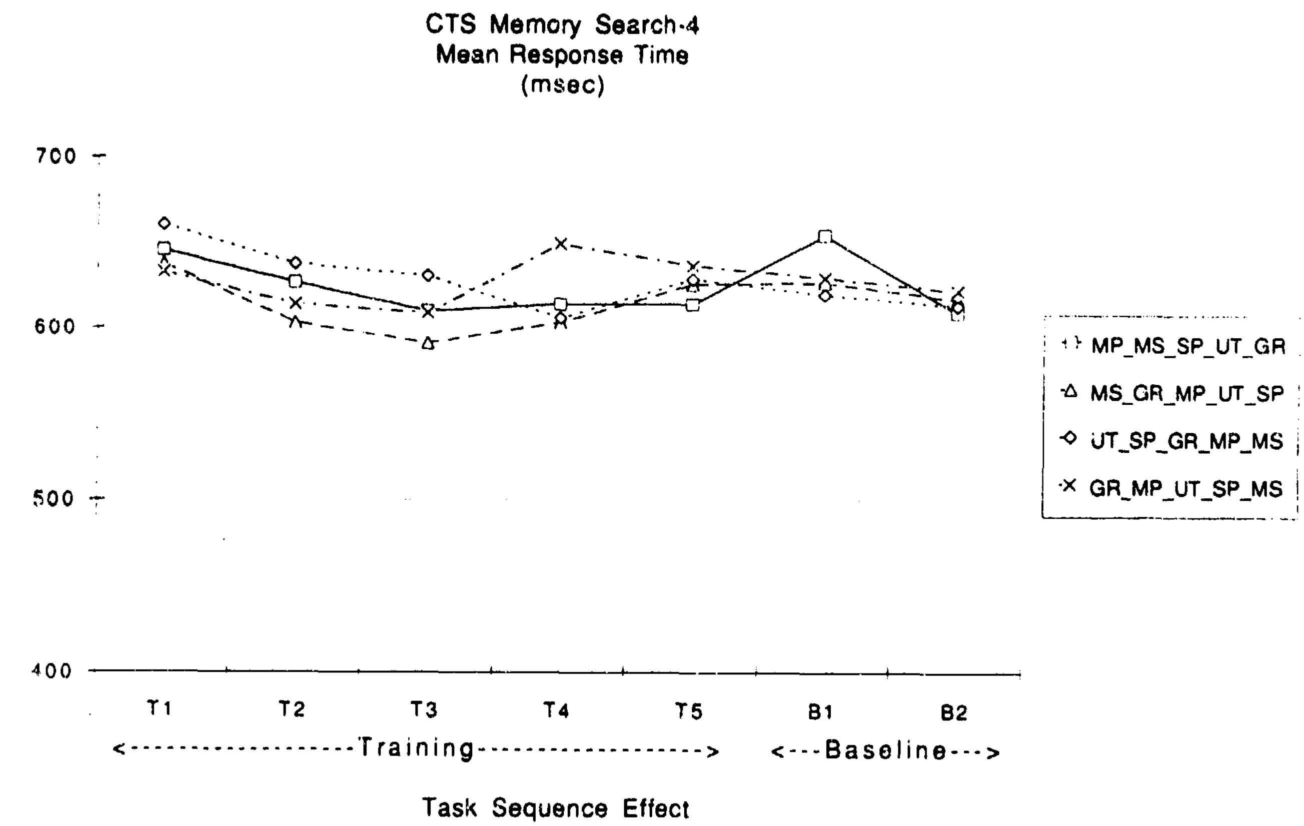
CTS Grammatical Reasoning
Percent Correct



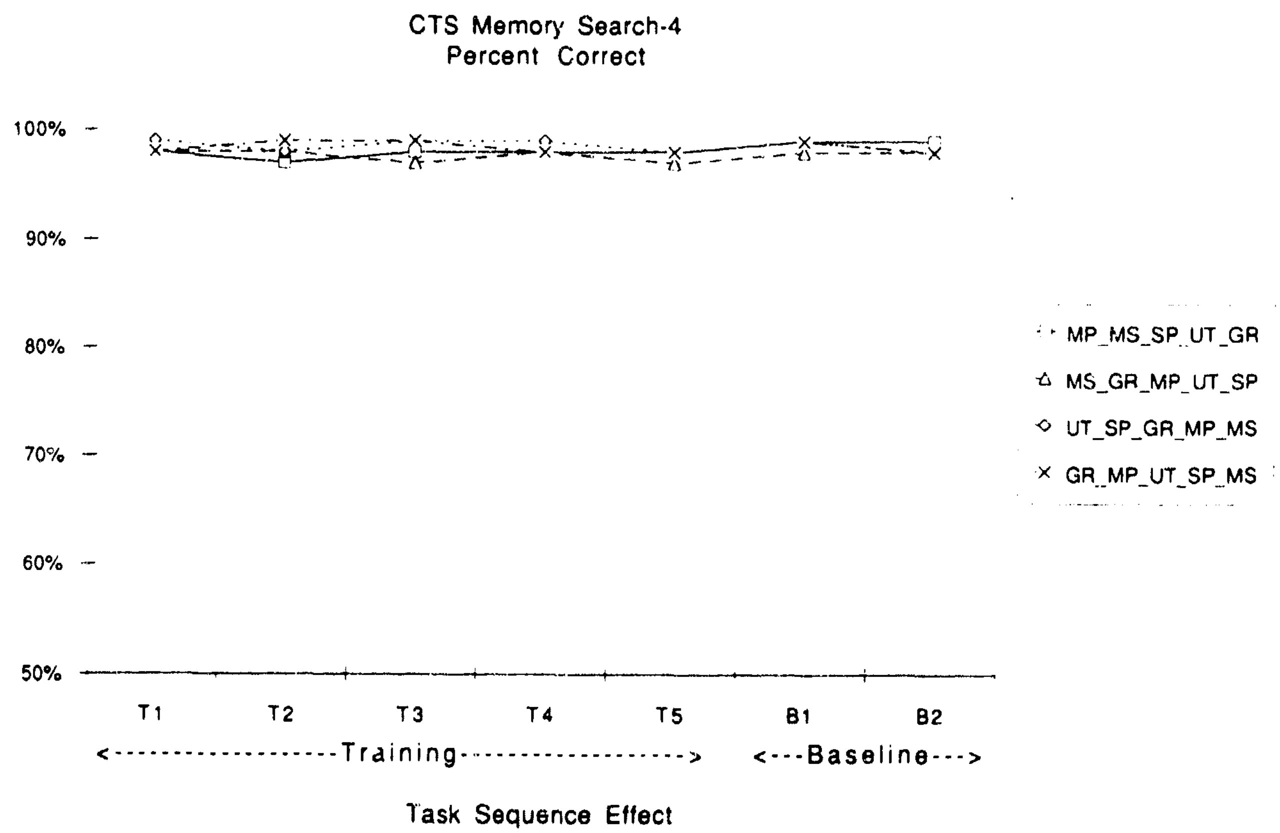
Task Sequence Effect


CTS Mathematical Processing
Mean Response Time
(msec)
2500
2000
1500
1000
500
0
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------> <---Baseline--->
Task Sequence Effect
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS

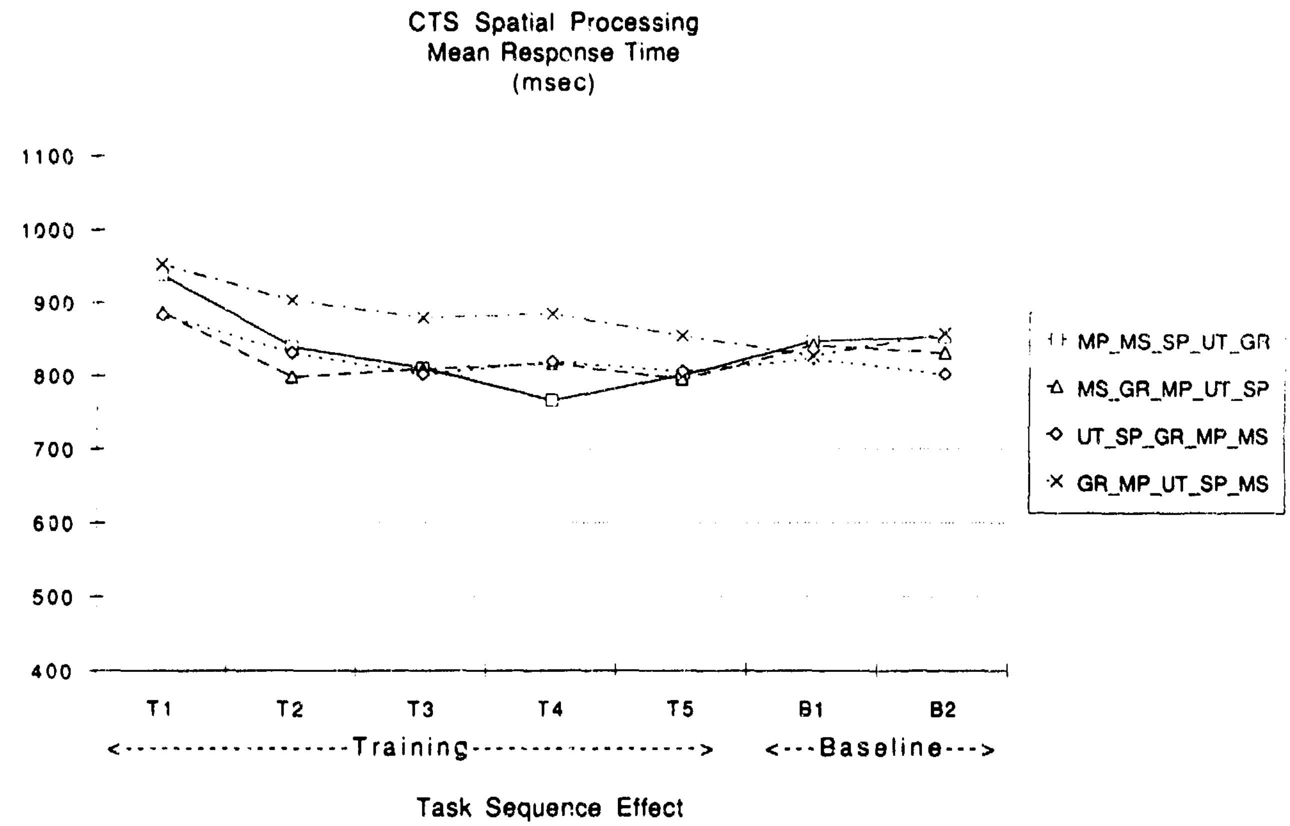

CTS Mathematical Processing
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------> <---Baseline--->
Task Sequence Effect
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS



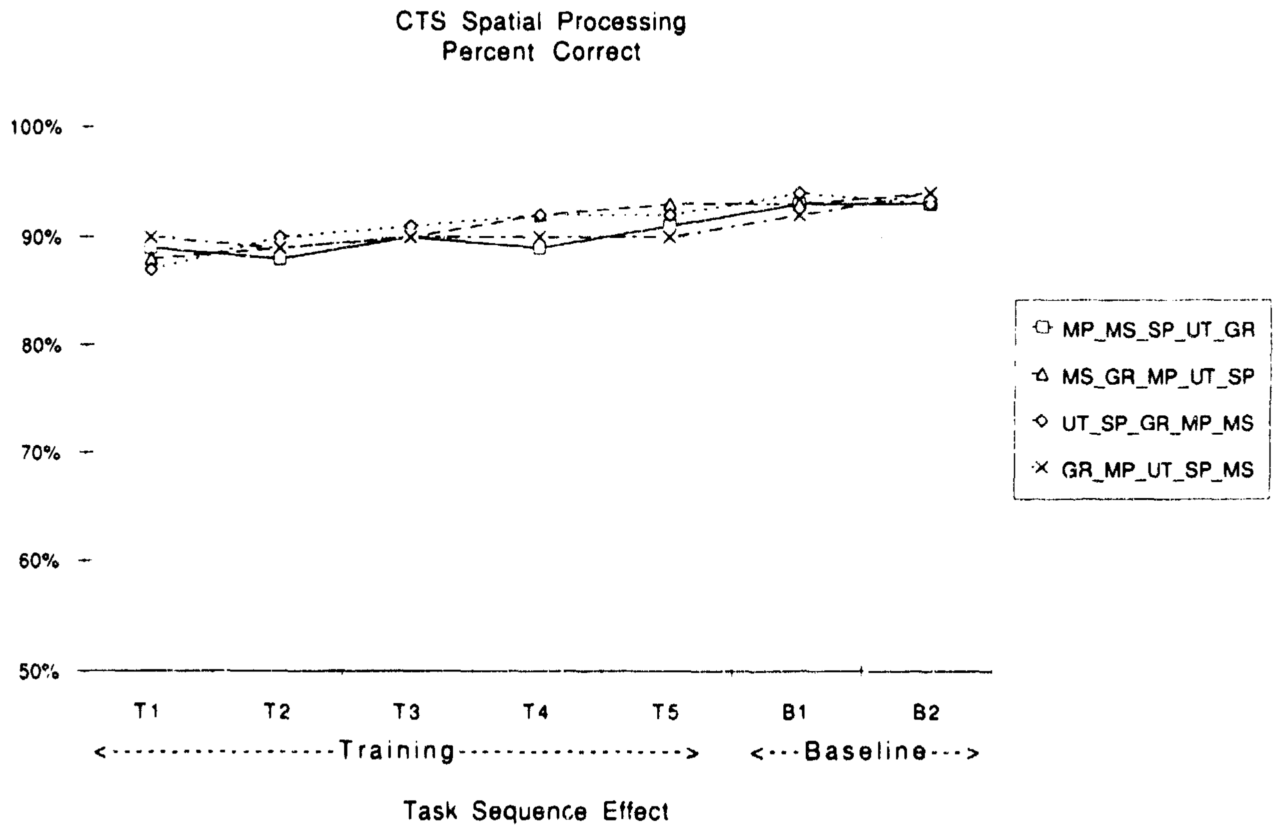
CTS Memory Search-4
Mean Response Time
(msec)
700
600
500
400
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
Task Sequence Effect



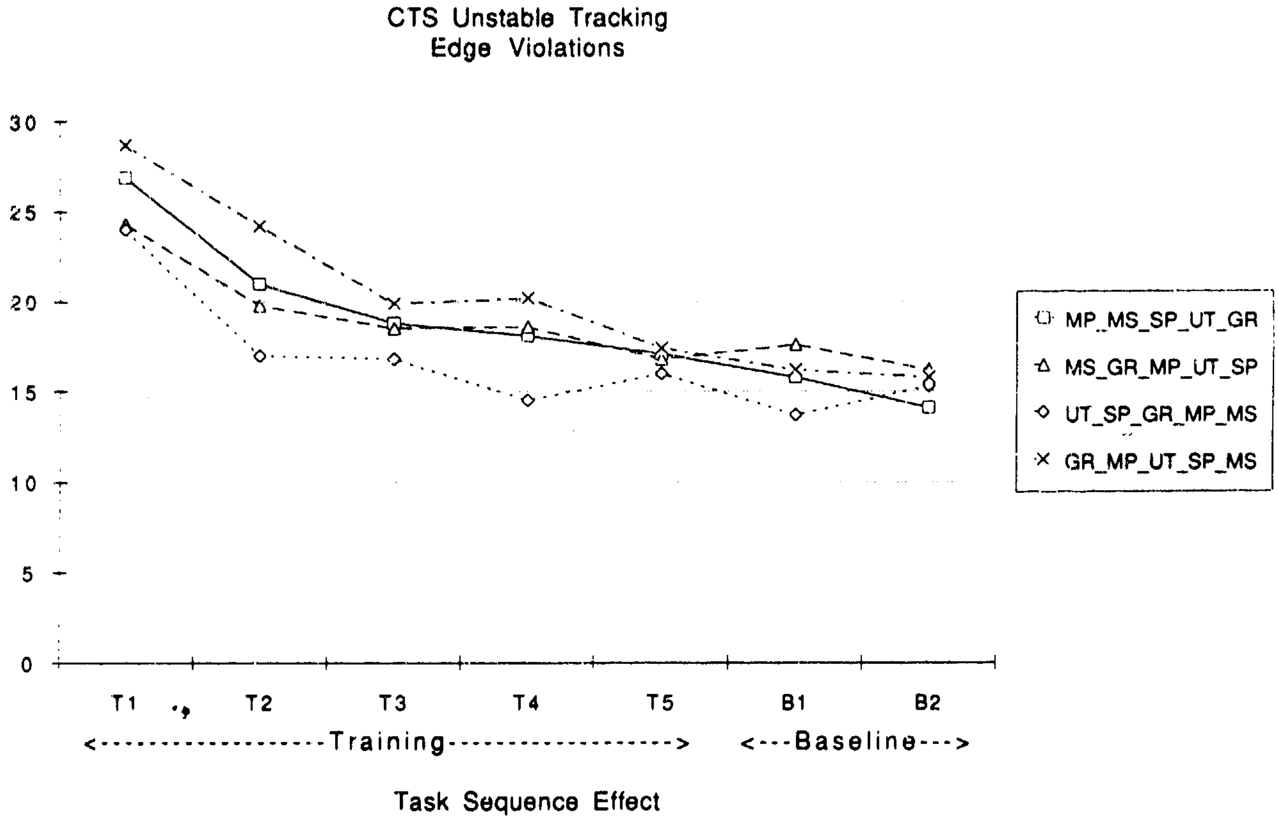
CTS Memory Search-4
Percent Correct
100%
90%
80%
70%
60%
50%
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
Task Sequence Effect

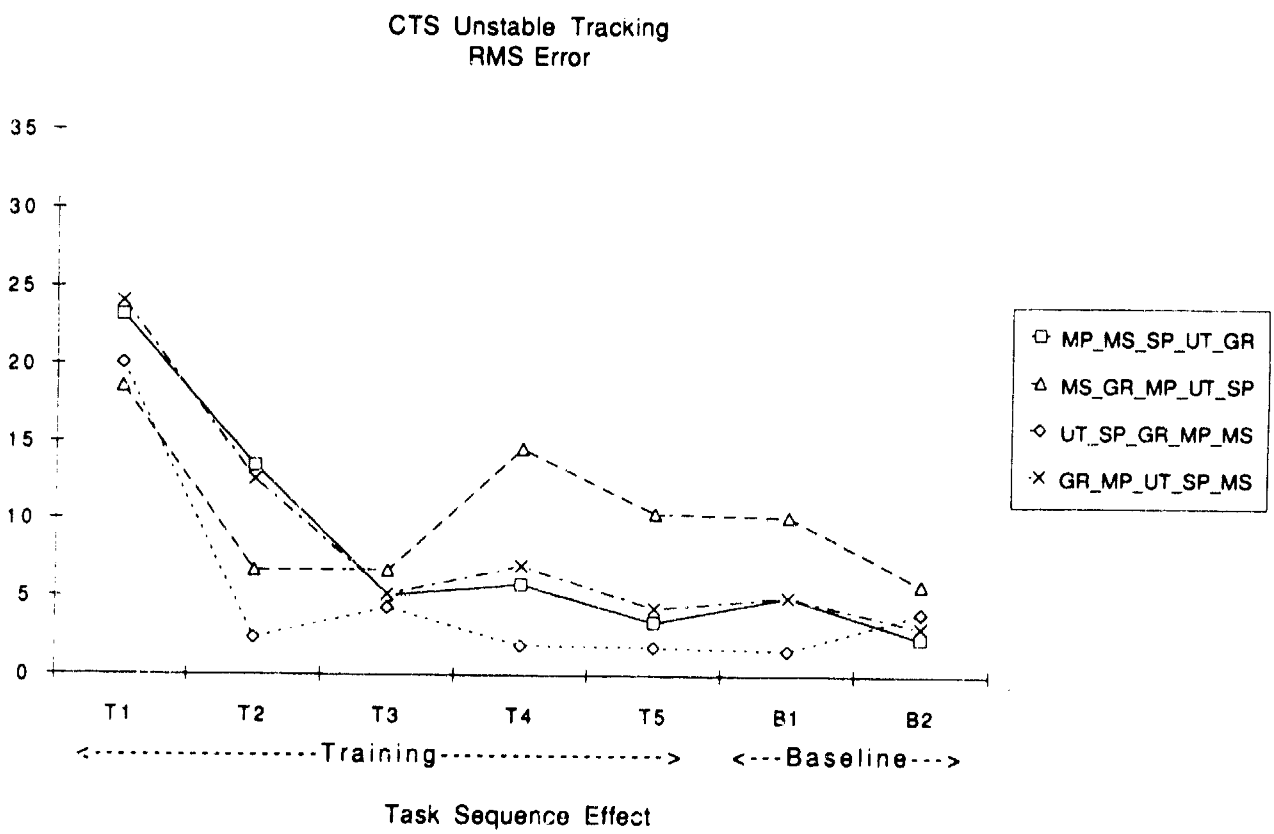


CTS Spatial Processing
Mean Response Time
(msec)
1100
1000
900
800
700
600
500
400
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
Task Sequence Effect
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS



CTS Spatial Processing
Percent Correct
100%
90%
80%
70%
60%
50%
T1
T2
T3
T4
T5
B1
B2
<------------------Training------------------>
<---Baseline--->
Task Sequence Effect
MP_MS_SP_UT_GR
MS_GR_MP_UT_SP
UT_SP_GR_MP_MS
GR_MP_UT_SP_MS





U.S.G.P.O.:1993-750-061/81018